

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 01:40:43 ; Search time 59.875 Seconds  
(without alignments)  
3923.374 Million cell updates/sec

Title: US-09-000-062-7

Perfect score: 494

Sequence: 1 CTCAGGCGAAGAACAGGTAT.....CGCAGATCCCGGATCGG 494

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/2CTUS COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length  | ID                  | Description        |
|------------|-------|-------------|---------|---------------------|--------------------|
| 1          | 494   | 100.0       | 494     | US-09-000-062-7     | Sequence 7, Appli  |
| 2          | 49.6  | 10.0        | 7218    | US-08-232-463-14    | Sequence 14, Appli |
| 3          | 38.2  | 7.7         | 6669    | US-10-204-708-5     | Sequence 5, Appli  |
| 4          | 37.2  | 7.5         | 832     | US-09-621-976-2813  | Sequence 2813, Ap  |
| 5          | 37.2  | 7.5         | 1230025 | US-09-198-452A-1    | Sequence 1, Appli  |
| 6          | 37.2  | 7.5         | 1193    | US-09-372-422A-23   | Sequence 23, Appli |
| 7          | 36.2  | 7.3         | 5562    | US-10-204-708-63    | Sequence 63, Appli |
| 8          | 36.2  | 7.3         | 6866    | US-10-204-708-20    | Sequence 20, Appli |
| 9          | 35.8  | 7.2         | 1493    | US-08-340-820-24    | Sequence 24, Appli |
| 10         | 35.8  | 7.2         | 1493    | US-08-593-535-24    | Sequence 24, Appli |
| 11         | 35.8  | 7.1         | 10467   | US-10-204-708-1     | Sequence 1, Appli  |
| 12         | 34.8  | 7.0         | 289     | US-09-007-708-17    | Sequence 17, Appli |
| 13         | 34.8  | 7.0         | 289     | US-09-244-796-17    | Sequence 17, Appli |
| 14         | 34.4  | 7.0         | 10467   | US-10-204-708-2     | Sequence 2, Appli  |
| 15         | 34.2  | 6.9         | 1406    | US-08-913-842-6     | Sequence 6, Appli  |
| 16         | 34    | 6.9         | 1664976 | US-08-916-421B-1    | Sequence 1, Appli  |
| 17         | 33.8  | 6.8         | 1051    | US-09-396-149-15    | Sequence 15, Appli |
| 18         | 33.6  | 6.8         | 531     | US-09-328-352-2881  | Sequence 2881, Ap  |
| 19         | 33.6  | 6.8         | 7304    | US-10-204-708-43    | Sequence 43, Appli |
| 20         | 33.6  | 6.8         | 11015   | US-10-204-708-55    | Sequence 55, Appli |
| 21         | 33.2  | 6.7         | 1896    | US-09-107-532A-894  | Sequence 894, Appl |
| 22         | 32.8  | 6.6         | 5666    | US-10-204-708-29    | Sequence 29, Appli |
| 23         | 32.8  | 6.6         | 8961    | US-10-204-708-80    | Sequence 80, Appli |
| 24         | 32.8  | 6.6         | 15071   | US-09-358-082A-29   | Sequence 29, Appli |
| 25         | 32.6  | 6.6         | 11014   | US-08-956-171E-91   | Sequence 91, Appli |
| 26         | 32.4  | 6.6         | 1545    | US-09-107-532A-2628 | Sequence 2628, Ap  |
| 27         | 32.4  | 6.6         | 5360    | US-10-204-708-65    | Sequence 65, Appli |

|    |      |     |       |   |                     |                    |
|----|------|-----|-------|---|---------------------|--------------------|
| 28 | 32.2 | 6.5 | 8537  | 4 | US-10-204-708-41    | Sequence 41, Appli |
| 29 | 32.2 | 6.5 | 8607  | 4 | US-10-204-708-71    | Sequence 71, Appli |
| 30 | 32   | 6.5 | 2022  | 4 | US-08-356-171E-244  | Sequence 244, App  |
| 31 | 32   | 6.5 | 46718 | 4 | US-09-816-093-3     | Sequence 3, Appli  |
| 32 | 31.8 | 6.4 | 476   | 4 | US-09-621-976-10903 | Sequence 10903, A  |
| 33 | 31.8 | 6.4 | 4998  | 4 | US-09-453-702B-21   | Sequence 21, Appli |
| 34 | 31.8 | 6.4 | 6070  | 4 | US-10-204-708-10    | Sequence 10, Appli |
| 35 | 31.8 | 6.4 | 6113  | 4 | US-10-204-708-14    | Sequence 14, Appli |
| 36 | 31.6 | 6.4 | 789   | 4 | US-09-134-001C-400  | Sequence 400, App  |
| 37 | 31.6 | 6.4 | 1453  | 4 | US-09-976-594-760   | Sequence 760, App  |
| 38 | 31.6 | 6.4 | 2301  | 4 | US-09-134-001C-852  | Sequence 852, App  |
| 39 | 31.6 | 6.4 | 4072  | 3 | US-09-272-496-7     | Sequence 7, Appli  |
| 40 | 31.6 | 6.4 | 11126 | 4 | US-08-956-171E-171  | Sequence 171, App  |
| 41 | 31.6 | 6.4 | 11131 | 4 | US-10-204-708-27    | Sequence 27, Appli |
| 42 | 31.6 | 6.4 | 41100 | 4 | US-09-755-665-46    | Sequence 46, Appli |
| 43 | 31.4 | 6.4 | 1884  | 1 | US-08-307-382-3     | Sequence 3, Appli  |
| 44 | 31.4 | 6.4 | 1884  | 1 | US-08-366-779-3     | Sequence 3, Appli  |
| 45 | 31.4 | 6.4 | 1884  | 1 | US-08-478-772-3     | Sequence 3, Appli  |

ALIGNMENTS

RESULT 1  
US-09-000-062-7  
; Sequence 7, Application US/09000062

; Patent No. 6338961  
; GENERAL INFORMATION:

; APPLICANT: DEROSE, Richard

; APPLICANT: CHAUBET, Nicole

; TITLE OF INVENTION: ISOLATED DNA SEQUENCE CAPABLE OF SERVING AS REGULATORY

; TITLE OF INVENTION: ELEMENT IN A CHIMERIC GENE WHICH CAN BE USED FOR THE

; TITLE OF INVENTION: TRANSFORMATION OF PLANTS

; FILE REFERENCE: 022650-453

; CURRENT APPLICATION NUMBER: US/09/000,062

; CURRENT FILING DATE: 1998-05-29

; EARLIER APPLICATION NUMBER: PCI/FR96/01109

; EARLIER FILING DATE: 1996-07-17

; EARLIER APPLICATION NUMBER: FR 95/08980

; EARLIER FILING DATE: 1995-07-19

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 7

; LENGTH: 494

; TYPE: DNA

; ORGANISM: Zea mays

; US-09-000-062-7

Query Match 100.0%; Score 494; DB 4; Length 494;

Best Local Similarity 100.0%; Pred. No. 2e-128;

Matches 494; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGGCGAAGAACAGGTATGTTTGTGTAATTAGATCAGGGTTTAGTCTTTCCAT 60

Db 1 CTCAGGCGAAGAACAGGTATGTTTGTGTAATTAGATCAGGGTTTAGTCTTTCCAT 60

QY 61 TACTTTTAAATGTTTTTCTGTTACTGTCCTCCGATCTGATTTTACGACAAATAGATTT 120

Db 61 TACTTTTAAATGTTTTTCTGTTACTGTCCTCCGATCTGATTTTACGACAAATAGATTT 120

QY 121 CGGGTTTTGTCCTCCATCCAGTTTGAAAAATAAAGTCGCTCTTTTAAAGTTTGCTGGATCGA 180

Db 121 CGGGTTTTGTCCTCCATCCAGTTTGAAAAATAAAGTCGCTCTTTTAAAGTTTGCTGGATCGA 180

QY 181 TAAACCTGTGAGATTTGAGTCTAGTCGATTTATTTGGATGATCCATTTTCTATCGTTTTT 240

Db 181 TAAACCTGTGAGATTTGAGTCTAGTCGATTTATTTGGATGATCCATTTTCTATCGTTTTT 240

QY 241 TCTTGTCTCGAAGTTCTGTATACAGATTTGTCGTGTGCGATTGTCAATACCTAGCCG 300

Db 241 TCTTGTCTCGAAGTTCTGTATACAGATTTGTCGTGTGCGATTGTCAATACCTAGCCG 300



```
QY 230 CATCGTTTCTTCTGCTTCGAGTCTGTATACCAAGATTGTCTG 276
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 5359 TGATTTTTCGATTAGTATTTAGTTAGTAGCGATATTATTG 5405

RESULT 4
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639083
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

Query Match 7.5%; Score 37.2; DB 4; Length 832;
Best Local Similarity 9.6%; Pred. No. 0.31;
Matches 35; Conservative 176; Mismatches 154; Indels 1; Gaps 1;

QY 71 TGTCTTTCTGTTACTGTCCTCCGAGTCTGATTTTACGACAATAGAGTTTCGGGTTTGT 130
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 YRWYWKYTTTAAKWTCKKWSWSYMYWYKWTYWRWRKKKAWKWTWTWY 60

QY 131 CCATTCACGTTGAAATAAAGTCGCTCTTTTAACTTTGCTGGATCGATAAACCTGTG 190
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 RYAMWGYKKKAMCMTKTKKKKGGYMMWYWGRRSYMAWTRTWGYAYRSMYWR 120

QY 191 AAGATTGAGTCTAGTCTGATTTATGGAATGATCCATCTTCATCGTTTCTTCTGCTCG 250
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 YRCWKXAYYKTTTCYSSKGTWWRKKAATTTWKKTYWAATRYMMWMCWKWRAS 180

QY 251 AAGTCTGTATAACCAATTTGCTGTGCGATTGTCATTACTAGCCGTGTATCGAGA 310
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 WYCWYWKARKWSTWRSRSYASARSASAKRCCYSCSWGMSWKYMMWRWRWATGAGM 240

QY 311 ACTAGGGTTTCGAGTCAAATTTCCCTCTTTGTTTATATCTGTTTCGATPACGATTTCAT 370
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 KAWRASCMRRKYAGSKTSYKSMCMWTRSKYCYTKARWTGYCYRKGMMGKGRWY 300

QY 371 -CTGATTAGGGTTTAAAGTGGGACGTTAGTATTCAATTTCTTCAAAATTAGTTAT 429
| | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 ASKYMWRKRWYRYSTGTGRASWWRWYTTMMKWKYAWARAARWMMWAW 360

QY 430 GGATAA 435
| | | | |
Db 361 RRACAA 366

RESULT 5
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifalbe, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
```





```

RESULT 6
US-09-372-422A-23/c
; Sequence 23, Application US/09372422A
; Patent No. 6313375
; GENERAL INFORMATION:
; APPLICANT: Rudolf Jung
; APPLICANT: Francois Barrieu
; TITLE OF INVENTION: Maize Aquaporins and Uses Thereof
; FILE REFERENCE: 0919
; CURRENT APPLICATION NUMBER: US/09/372,422A
; CURRENT FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: US 60/098,692
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 23
; LENGTH: 1193
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (88)...(838)
;
US-09-372-422A-23

```

```

QY      78   TCCTGTTACTCTCTCCCGCATCTGATTTTACGACAATAAGATTTCGGGTTCGTTTGTCCTAATC 137
DB      1122 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAAACGAAAAATCACTGTTTTTCCCAC 1063

QY      138   CAGTTTGAAAATAAAACGTCGC 158
DB      1062 CAAGCGCATCGAAAACGGCAG 1042

RESULT 7
US-10-204-708-63
; Sequence 63, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; PRIORITY FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIORITY FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIORITY FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIORITY FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIORITY FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIORITY FILING DATE: 2000-06-30

```

```

; NUMBER OF SEQ ID NOS: 98
;
; SEQ ID NO 63
; LENGTH: 5562
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

```







```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 07-073043
; FILING DATE: 30-MAR-1995
; PRIOR APPLICATION DATA: PCT/JP96/00777
; APPLICATION NUMBER: PCT/JP96/00777
; FILING DATE: 26-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: OHBA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1406 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-913-842-6

Query Match      6.9%; Score 34.2; DB 3; Length 1406;
Best Local Similarity 53.3%; Pred. No. 2.5;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 339 TTTGGTTATATCTCGTTCCGATACGATTCACTCGGATTAGGTTTAAAGTGGTGACGTT 398
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1118 TATTATTATATTTTATTACAAATAATTTCTAATAGATATCTAATCTCCTTTGAGACATT 1059

Qy 399 TAGTATTCCAATTTCTTCAAAATTTAGTTATGGATAATGAAATCCCGAATTGACTGTC 458
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1058 TTTGGTTTATTTTATATAATTTTAAATTTCTTTTATAAATAATGACTCTTCGTTG 999

Qy 459 AATTCTTTGTTAAAT 473
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
998 ATTCAAAATTTAAAT 984
```

Search completed: June 20, 2004, 05:03:44  
Job time : 73.875 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 20, 2004, 00:55:23 ; Search time 2930.58 Seconds  
(without alignments)  
6182.178 Million cell updates/sec

Title: US-09-000-062-6  
Perfect score: 418  
Sequence: 1 tgaagtagattcttcgattc.....taattgtgaacagatccc 418

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rtd.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID        | Description        |
|------------|-------|---------------|--------|----|-----------|--------------------|
| 1          | 418   | 100.0         | 418    | 6  | A59349    | A59349 Sequence 6  |
| 2          | 418   | 100.0         | 418    | 6  | AR182671  | AR182671 Sequence  |
| 3          | 415   | 99.3          | 4833   | 8  | ATH3G     | X60429 A.thaliana  |
| C 4        | 411.8 | 98.5          | 112067 | 8  | ATCHRIV92 | AL161596 Arabidops |
| C 5        | 411.8 | 98.5          | 112067 | 8  | ATTSJ17   | AL035708 Arabidops |
| 6          | 54.4  | 13.0          | 5998   | 6  | AX34536   | AX34536 Sequence   |
| 7          | 52    | 12.4          | 349980 | 6  | AX344561  | AX344561 Sequence  |
| 8          | 51.2  | 12.2          | 253151 | 3  | AE014842  | AE014842 Plasmodiu |
| 9          | 50.6  | 12.1          | 17527  | 6  | AX339160  | AX339160 Sequence  |
| 10         | 50.6  | 12.1          | 17527  | 6  | AX346334  | AX346334 Sequence  |
| 11         | 50.2  | 12.0          | 6306   | 6  | AX278060  | AX278060 Sequence  |
| 12         | 50.2  | 12.0          | 6306   | 6  | AX323843  | AX323843 Sequence  |
| C 13       | 49.6  | 11.9          | 206653 | 2  | EX470257  | EX470257 Danio rer |
| 14         | 49.2  | 11.8          | 127172 | 2  | AC141780  | AC141780 Apis mell |
| 15         | 49.2  | 11.8          | 349980 | 6  | AX344573  | AX344573 Sequence  |
| 16         | 49.2  | 11.8          | 349980 | 6  | AX344574  | AX344574 Sequence  |
| 17         | 49    | 11.7          | 5987   | 6  | AX346464  | AX346464 Sequence  |
| 18         | 49    | 11.7          | 5987   | 6  | AX458585  | AX458585 Sequence  |
| 19         | 49    | 11.7          | 5987   | 6  | AX795831  | AX795831 Sequence  |
| 20         | 49    | 11.7          | 5987   | 6  | AX795947  | AX795947 Sequence  |
| 21         | 49    | 11.7          | 5987   | 6  | AX822343  | AX822343 Sequence  |
| 22         | 49    | 11.7          | 5987   | 6  | AX822471  | AX822471 Sequence  |
| 23         | 49    | 11.7          | 5987   | 6  | AX825983  | AX825983 Sequence  |
| 24         | 49    | 11.7          | 5987   | 6  | AX826111  | AX826111 Sequence  |
| 25         | 48.8  | 11.7          | 15479  | 6  | AX348350  | AX348350 Sequence  |
| 26         | 48.4  | 11.6          | 17848  | 6  | AX277865  | AX277865 Dictyoste |
| 27         | 48.4  | 11.6          | 17848  | 6  | AX277865  | AX277865 Sequence  |
| 28         | 48.4  | 11.6          | 17848  | 6  | AX323550  | AX323550 Sequence  |
| 29         | 48.4  | 11.6          | 17848  | 6  | AX348363  | AX348363 Sequence  |
| 30         | 48    | 11.5          | 2501   | 6  | AX598874  | AX598874 Sequence  |
| 31         | 48    | 11.5          | 2501   | 6  | AX599020  | AX599020 Sequence  |
| 32         | 48    | 11.5          | 6059   | 6  | AX346382  | AX346382 Sequence  |
| 33         | 48    | 11.5          | 6245   | 6  | AX251202  | AX251202 Sequence  |
| 34         | 48    | 11.5          | 16724  | 6  | AX281425  | AX281425 Sequence  |
| 35         | 48    | 11.5          | 16724  | 6  | AX345992  | AX345992 Sequence  |
| 36         | 48    | 11.5          | 16724  | 6  | AX348591  | AX348591 Sequence  |
| C 37       | 48    | 11.5          | 28568  | 3  | AC116955  | AC116955 Dictyoste |
| 38         | 48    | 11.5          | 136791 | 9  | AC079896  | AC079896 Homo sapi |
| 39         | 47.8  | 11.4          | 349980 | 6  | AX344565  | AX344565 Sequence  |
| 40         | 47.6  | 11.4          | 5649   | 6  | AX767532  | AX767532 Sequence  |
| 41         | 47.6  | 11.4          | 5649   | 6  | AX795900  | AX795900 Sequence  |
| 42         | 47.6  | 11.4          | 5649   | 6  | AX822418  | AX822418 Sequence  |
| 43         | 47.6  | 11.4          | 5649   | 6  | AX826058  | AX826058 Sequence  |
| 44         | 47.4  | 11.3          | 21537  | 6  | AX346901  | AX346901 Sequence  |
| 45         | 47.4  | 11.3          | 349980 | 6  | AX344563  | AX344563 Sequence  |

ALIGNMENTS

RESULT 1  
A59349 A59349 418 bp DNA linear PAT 06-MAR-1998  
LOCUS Sequence 6 from Patent WO9704114.  
DEFINITION A59349  
ACCESSION A59349  
VERSION A59349.1 GI:3714675  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE 1  
AUTHORS Derose,R., Chaubet,N. and Gigot,C.  
TITLE ISOLATED DNA SEQUENCE FOR USE AS A REGULATOR REGION IN A CHIMERIC

GENE USEFUL FOR TRANSFORMING PLANTS  
Patent: WO 9704114-A 6 06-FEB-1997;  
RHONE POULENC AGROCHIMIE (FR)  
Other publication FR 2736929 970124.  
Location/Qualifiers

1. .418  
/organism="Arabidopsis thaliana"  
/mol\_type="unassigned DNA"  
/strain="COLUMBIA"  
/db\_xref="taxon:3702"  
/clone="C22"  
/clone\_lib="COSMID"

ORIGIN

Query Match 100.0%; Score 418; DB 6; Length 418;  
Best Local Similarity 100.0%; Pred. No. 2.le-81;  
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGGTACGATCTTCGATCCCTTTGATTTCTCGGAATATTTTTCGGTGATCGTA 60  
DB 1 TGAGGTACGATCTTCGATCCCTTTGATTTCTCGGAATATTTTTCGGTGATCGTA 60  
QY 61 AACTACTGGAATCGCTCGATAGGTGATGAGAAATTTAGCGGAGATTAGTTCTATTCTTGG 120  
DB 61 AACTACTGGAATCGCTCGATAGGTGATGAGAAATTTAGCGGAGATTAGTTCTATTCTTGG 120  
QY 121 CCATTATCTTGTCTTCGCGGAATGATCTCCGATATAAGATTTAGGTTAGAGATGAA 180  
DB 121 CCATTATCTTGTCTTCGCGGAATGATCTCCGATATAAGATTTAGGTTAGAGATGAA 180  
QY 181 TCGTATAGTAGATTTTCATCCAGATAGTTTCTTTGTCTAGAAATCTCTGAAATCTTCGA 240  
DB 181 TCGTATAGTAGATTTTCATCCAGATAGTTTCTTTGTCTAGAAATCTCTGAAATCTTCGA 240  
QY 241 TAGTTTTCACATGCTGTAATAGATTTCTTATTCGGCGATTGTTGATAGGGTTTTCAT 300  
DB 241 TAGTTTTCACATGCTGTAATAGATTTCTTATTCGGCGATTGTTGATAGGGTTTTCAT 300  
QY 301 TTTCTTGATTTGCGATTCGAATAGGATTTCTTTGGTTTGTGATCTTACGATA 360  
DB 301 TTTCTTGATTTGCGATTCGAATAGGATTTCTTTGGTTTGTGATCTTACGATA 360  
QY 361 CATTCCTGCAATGTAATAGTATGATCTAAATCTTTGTTAAATTTGTTGAACAGATCCC 418  
DB 361 CATTCCTGCAATGTAATAGTATGATCTAAATCTTTGTTAAATTTGTTGAACAGATCCC 418

RESULT 2

AR182671  
LOCUS 418 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 6 from patent US 638961.  
ACCESSION AR182671  
VERSION AR182671.1 GI:20225878  
KEYWORDS  
ORGANISM Unknown.  
SOURCE Unknown.  
UNCLASSIFIED.

REFERENCE 1 (bases 1 to 418)  
AUTHORS DeRose, R., Chaubet, N. and Gigot, C.  
TITLE Isolated DNA sequence capable of serving as regulatory element in a chimeric gene which can be used for the transformation of plants  
JOURNAL Patent: US 638961-A 6 13-JAN-2002;  
FEATURES Location/Qualifiers  
1. .418  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 418; DB 6; Length 418;  
Best Local Similarity 100.0%; Pred. No. 2.le-81;  
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAGGTACGATCTTCGATCCCTTTGATTTCTCGGAATATTTTTCGGTGATCGTA 60

DB 1 TGAGGTACGATCTTCGATCCCTTTGATTTCTCGGAATATTTTTCGGTGATCGTA 60  
QY 61 AACTACTGGAATCGCTCGATAGGTGATGAGAAATTTAGCGGAGATTAGTTCTATTCTTGG 120  
DB 61 AACTACTGGAATCGCTCGATAGGTGATGAGAAATTTAGCGGAGATTAGTTCTATTCTTGG 120  
QY 121 CCATTATCTTGTCTTCGCGGAATGATCTCCGATATAAGATTTAGGTTAGAGATGAA 180  
DB 121 CCATTATCTTGTCTTCGCGGAATGATCTCCGATATAAGATTTAGGTTAGAGATGAA 180  
QY 181 TCGTATAGTAGATTTTCATCCAGATAGTTTCTTTGTCTAGAAATCTCTGAAATCTTCGA 240  
DB 181 TCGTATAGTAGATTTTCATCCAGATAGTTTCTTTGTCTAGAAATCTCTGAAATCTTCGA 240  
QY 241 TAGTTTTCACATGCTGTAATAGATTTCTTATTCGGCGATTGTTGATAGGGTTTTCAT 300  
DB 241 TAGTTTTCACATGCTGTAATAGATTTCTTATTCGGCGATTGTTGATAGGGTTTTCAT 300  
QY 301 TTTCTTGATTTGCGATTCGAATAGGATTTCTTTGGTTTGTGATCTTACGATA 360  
DB 301 TTTCTTGATTTGCGATTCGAATAGGATTTCTTTGGTTTGTGATCTTACGATA 360  
QY 361 CATTCCTGCAATGTAATAGTATGATCTAAATCTTTGTTAAATTTGTTGAACAGATCCC 418  
DB 361 CATTCCTGCAATGTAATAGTATGATCTAAATCTTTGTTAAATTTGTTGAACAGATCCC 418

RESULT 3

ATH3G  
LOCUS 4833 bp DNA linear PLN 08-JAN-1996  
DEFINITION A.thaliana H3 gene 1 and H3 gene 2 for H3.3-like histone variant.  
ACCESSION X60429  
VERSION X60429.1 GI:16323  
KEYWORDS histone; histone H3; histone H3.3 homologue.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 4833)  
AUTHORS Chaubet, N., Clement, B. and Gigot, C.  
TITLE Genes encoding a histone H3.3-like variant in Arabidopsis contain intervening sequences  
JOURNAL J. Mol. Biol. 225 (2), 569-574 (1992)  
MEDLINE 92277663  
PUBMED 1593639  
REFERENCE 2 (bases 1 to 4833)  
AUTHORS Gigot, C.  
TITLE Direct Submission  
JOURNAL Submitted (17-SEP-1991) C. Gigot, Inst de Biologie Mol des Plantes, 12 Rue du General Zimmer, 67084 Strasbourg Cedex, FRANCE

FEATURES

1. .4833  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/cell\_line="Columbia"  
/clone\_lib="cosmid c22"  
/dev\_stage="callus"  
691..1100  
/gene="H3, gene1"  
691..1100  
/gene="H3, gene1"  
692..1100  
/number=1  
/gene="H3, gene1"  
/number=1  
/gene="H3, gene1"  
join(1101..1172,1260..1336,1431..1521,1620..1790)  
/gene="H3.gene1"  
join(1101..1172,1260..1336,1431..1521,1620..1790)  
/gene="H3.gene1"



/codon\_start=1  
/product="histone H3.3 like protein"  
/protein\_id="CAA42957.1"  
/db\_xref="GI:16324"  
/db\_xref="GOA:P59169"  
/db\_xref="SWISS-PROT:P59169"  
/translation="MARTKQARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP  
GTVALREIKYOKSTELLIRKLPPORLVREIAQDFKTDLRFSQHAVLALQEAAYLV  
GLFEDTNLCATHAKRVITMPKIQIARRIGERA"  
1101..1172  
/gene="H3, gene1"  
1101..1172  
/exon  
/gene="H3, gene1"  
/number=2  
1173..1259  
/gene="H3, gene1"  
1173..1259  
/gene="H3, gene1"  
/number=2  
1260..1336  
/gene="H3, gene1"  
1260..1336  
/gene="H3, gene1"  
/number=3  
1337..1430  
/gene="H3, gene1"  
1337..1430  
/gene="H3, gene1"  
/number=3  
1431..1521  
/gene="H3, gene1"  
1431..1521  
/gene="H3, gene1"  
/number=4  
1522..1619  
/gene="H3, gene1"  
1522..1619  
/gene="H3, gene1"  
1620..1790  
/gene="H3, gene1"  
1620..1790  
/gene="H3, gene1"  
/number=5  
join(2912..2983,3065..3141,3227..3488)  
/gene="H3, gene2"  
join(2912..2983,3065..3141,3227..3488)  
/gene="H3, gene2"  
/codon\_start=1  
/product="histone H3.3 like protein"  
/protein\_id="CAA42958.1"  
/db\_xref="GI:404825"  
/db\_xref="GOA:P59169"  
/db\_xref="SWISS-PROT:P59169"  
/translation="MARTKQARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP  
GTVALREIKYOKSTELLIRKLPPORLVREIAQDFKTDLRFSQHAVLALQEAAYLV  
GLFEDTNLCATHAKRVITMPKIQIARRIGERA"  
2912..2983  
/gene="H3, gene 2"  
2912..2983  
/exon  
/gene="H3, gene 2"  
/number=1  
2984..3064  
/gene="H3, gene 2"  
2984..3064  
/gene="H3, gene 2"  
/number=1  
3065..3141  
/gene="H3, gene 2"  
3065..3141  
/gene="H3, gene 2"  
/number=2  
3142..3226  
/gene

intron  
/gene="H3, gene 2"  
3142..3226  
/gene="H3, gene 2"  
/number=2  
3227..3488  
/gene="H3, gene 2"  
3227..3488  
/gene="H3, gene 2"  
/number=3  
ORIGIN  
Query Match 99.3%; Score 415; DB 8; Length 4833;  
Best Local Similarity 100.0%; Pred. No. 7.1e-81;  
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAGGTACGATCTTCGATCCCTCTTTGATTTTCCTCGAAATATTTTTCGGTGATCGTA 60  
Db |||||  
QY 61 AACTACTGGAATCGTCGATAGGTGCTAGCAATATAGGCGAGATTAGTTCTTATCTTGG 120  
Db |||||  
QY 748 AACTACTGGAATCGTCGATAGGTGCTAGCAATATAGGCGAGATTAGTTCTTATCTTGG 807  
Db |||||  
QY 121 CCATTATCTGTTTCTTCGCGGAATGATCTTCGATATAAGATTTTAGGTTAGAGATGAA 180  
Db |||||  
QY 808 CCATTATCTGTTTCTTCGCGGAATGATCTTCGATATAAGATTTTAGGTTAGAGATGAA 867  
Db |||||  
QY 181 TCGTATAGCTAGATTTTCATCACCAGATAGTTCTTTGTCTAGAAATCTCGAAATCTCGA 240  
Db |||||  
QY 868 TCGTATAGCTAGATTTTCATCACCAGATAGTTCTTTGTCTAGAAATCTCGAAATCTCGA 927  
Db |||||  
QY 241 TAGTTTTCACATGTTAAATAGATTTCTTATTCGCGGATTTCTTTGTTGTTGTTGTTGAT 300  
Db |||||  
QY 928 TAGTTTTCACATGTTAAATAGATTTCTTATTCGCGGATTTCTTTGTTGTTGTTGAT 987  
Db |||||  
QY 301 TTTCTTGATTATGCGATTGCAATTAGGGAATTTCTTTGTTGTTGTTGTTGTTGTTGAT 360  
Db |||||  
QY 988 TTTCTTGATTATGCGATTGCAATTAGGGAATTTCTTTGTTGTTGTTGTTGTTGAT 1047  
Db |||||  
QY 361 CATTCCTGCAATGATACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 415  
Db |||||  
QY 1048 CATTCCTGCAATGATACATGATGATGATGATGATGATGATGATGATGATGATGAT 1102  
Db |||||

RESULT 4  
ATCHRIV92/c  
LOCUS Arabidopsis thaliana DNA chromosome 4, contig fragment No. 92.  
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 92.  
ACCESSION AL161596  
VERSION AL161596.2 GI:7271037  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
1 (bases 1 to 107700)  
Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and  
Mayer,K.F.X.  
Unpublished  
2 (bases 9546 to 9977)  
Voickaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W.,  
Lemcke,K. and Mayer,K.F.X.  
Unpublished  
3 (bases 107578 to 112067)  
Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and  
Mayer,K.F.X.  
Unpublished  
4 (bases 1 to 112067)  
EU Arabidopsis sequencing project.  
Direct Submission  
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer  
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:

lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project  
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge  
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
 E-mail: michael.bevan@bbsrc.ac.uk  
 Information on performance of analysis and a more detailed  
 annotation of this entry and other sequences of chromosomes 3, 4  
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>  
 this fragment has an overlap with ATCHRIV91 at the 5' end and an  
 overlap with ATCHRIV93 at the 3' end.

## COMMENT

## FEATURES

```

source
  1..11267
  /organism="Arabidopsis thaliana"
  /mol_type="genomic DNA"
  /variety="Columbia"
  /db_xref="taxon:3702"
  /chromosome="4"
  6122..8203
  /gene="AT4g39830"
  join(6122..6415,6481..6573,6652..6864,6959..7153,
  7250..8203)
  /note="strong similarity to L-ascorbate oxidase, Cucumis
  sativus, PIR1:KSKVAO
  Contains Glycosyl hydrolases family 5 signature
  AA384-393; Multicopper oxidases signatures
  AA131-151; Multicopper oxidases signatures
  AA540-540; Multicopper oxidases signatures
  AA543-556; Prokaryotic membrane lipoprotein lipid
  attachment site AA567-577
  contains EST gb:A1994215.1, AA585817, A1996008.1"
  /codon_start=1
  /product="putative L-ascorbate oxidase"
  /protein_id="CAB80646.1"
  /db_xref="GI:7271038"
  /db_xref="GOA:O65670"
  /db_xref="SPTREMBL:O65670"
  /translation="MKRPKSDSDTVHVNLMVLFIALPFSVLCQKIRRFKEVY
  EFKSPDCEKLVITINGPFGPTDKAQQDITVVELKNSNTENVAVHHGIRQITGP
  WFDVGVGVYFVIFQVVDREGTYHSHYQMORSEGLIGMIQVSPATPEP
  EFTYDVRNFIITDTHYKSKATGLASIPFKWVGEPQSLMTQGRFNCNNLTT
  PSLVSGVGNVSNADCSRFLITVI PKTYRLRIQSLTALSALSPOIGHNLTVEADG
  HVVEPTVKNLVSGEYTSVLLKADQNPENYITSSIVSRPATTPTATAVLYNEN
  HRRRPPTSESSNIPEWNDVRSLSLAQSLAKSRGRTHALPENSQKVIILLATONEV
  NRYRNSVNNVSHHPKPYLLALKQNTNAPDRPTAPENYDSRNYDIFAKPLNANA
  ITSDGIRFNSTVDVILQNTNANNSETHPWHLGHDFVWLGEGKFNESDPP
  KRYNRVDP1KKNTVAVQPFQGTAFURFRADNPGVWSFCHIESHFHMGVIFSSGIDK
  VSSLPSISNGCCQTKR"
  6122..6415
  /gene="AT4g39830"
  /number=1
  6416..6480
  /gene="AT4g39830"
  /number=1
  6481..6573
  /gene="AT4g39830"
  /number=2
  6574..6651
  /gene="AT4g39830"
  /number=2
  6652..6864
  /gene="AT4g39830"
  /number=3
  6865..6958
  /gene="AT4g39830"
  /number=3
  6959..7153
  /gene="AT4g39830"
  /number=4
  7154..7249
  /gene="AT4g39830"
  /number=4
  7250..8203
  /gene="AT4g39830"
  /number=5

gene
  12287..13642
  /gene="AT4g39840"
  12287..13642
  /gene="AT4g39840"
  /note="contains EST gb:A1999725.1, 246537, F13826,
  A1996686.1"
  /codon_start=1
  /product="putative protein"
  /protein_id="CAB80647.1"
  /db_xref="GI:7271039"
  /db_xref="SPTREMBL:Q9SMR9"
  /translation="MSPFKYLCFRKVLILLYSSLSLFFSSSSFFASVDDHQLRIDP
  HFVRRLVKDLDSDGDDDETNPFPFKKLLTDSVSSSSSSGFKQKQTKLLKPKIS
  SSSSTKQTKLAKTTMTGSHKLNSTKSNNTTKTSSELKLNSTGKSNSTSSIKKS
  ADLSKSSSSKNTTKIPKSSKLSPPSEKSKQSPKVPKSKOSEKEIKFPWLDDDED
  EDFVSERDLPTAFORSLLIPDLERISTSKNYINKANKOITKNFKFYGNKVAPTIAS
  VVSFVILVPLLVLSLIFNFKAYESLQKTLIIQIYLSIYSEILCLSLVATGIBPLK
  FLYATSSSTVYCLQIQLTQGVYFVYLLLMYLVLFSTDCGLKGLKVLGAQFVGPVAF
  GLHYIVAFHRVLDVRPKTNKWNKGVYATCFLLILLNNAERRKXEYLEEGDEGKK
  N"
  12287..13642
  /gene="AT4g39840"
  /number=1
  15539..16205
  /note="EST N95877 matches to positions 26106-26355,
  26723-26772; possibly related to subsequent gen"
  16239..23312
  /gene="AT4g39850"
  join(16239..16300,16394..16679,16981..17205,17341..17538,
  17675..18129,18222..18357,18433..18511,19039..19271,
  19326..19630,19712..19844,19901..20056,20213..20435,
  20515..20623,20845..20935,21205..21372,21453..21568,
  21626..21860,21924..22104,22162..22362,22534..22680,
  22750..22983,23134..23312)
  /gene="AT4g39850"
  /note="similarity to 70K peroxisomal membrane protein,
  Rattus norvegicus , PIR2:A35723
  Contains ABC transporters family signature AA580-594; ABC
  transporters family signature AA1369-1383; ATP/GTP-binding
  site motif A (p-loop) AA1247-1254
  contains EST gb:N95877"
  /codon_start=1
  /product="putative protein"
  /protein_id="CAB80648.1"
  /db_xref="GI:7271040"
  /db_xref="GOA:Q9SMR8"
  /db_xref="SPTREMBL:Q9SMR8"
  /translation="MPSLQLQLTERGRGLVASRRKSIILAAIGIYAAGGTAVYKSRV
  ASRRPSSRLCNGOSDDDETELEKLTADONAKITTKKGGGLKSLQVLTALLSOMG
  KMGARDLLAVATVVFPTALSNLAKVQGFLEPAAPLEAPLRLPFLISENIMLCFMLS
  TLHSTSYIYGALSIRPKKILKLIHSHYFENNMYIKISHVDGRIITHPQRIASDVPR
  FSSLSLIIDDLTAVTDGILYAWRUCSVASPKYIFWILAYVLGAGTAINRNPSPFGK
  LMSKEQLEGEYQRLHSLRTHSESAFYGETREESHIQCKPKNLVSHMSHVLHDPHW
  WFGMIQDFLLKLGATVAVILIIPEFSGHLRPPDSTLGRAMELNIRYTHVLSLF
  QALGTLISRRNLNLSGADRIHELMVSRSLSGDSSKSFQNRNRSYLSANVVEF
  SDVKVAVTGNVLVDLITLAVEQSNLLITGGLWPLVSGHIVKPGVSGDLNKEIYVPD
  ORYVAVGTLDLIYPLTQSGSELLTEIGWELLKNWSILVSLYFTWDLVLEIYLD
  RYQPEKVNKGDLSLGEQQLGWARLFVHKPKFALIDECTSAVTTDMERFAKTRA
  NGTSCITISHRPALVADHDVLSDEGSGSHYKRDSDALLTDAIDSXKSDTRQ
  NDAMVQVAFAPAAKVAEVLFSATNSKASQITQILARSPVDSKSVLFPFPQPT
  SQRALPRVAMNLVLPWKPHLLFCISTWLTWYSSCADANIDILLIKTHDFTVFOI
  PTIFDQACQLLAVACLVNRTLISDRISLNGTTVKYVLEQKAAFLRIGLSVLQS
  GASSIIAPSLRHLTORLALGWRIRLTQHLRLNRYNNAFYKVFHMSGNSIDADQRLTR
  DLEKLTADLSGLLTGMVKVSDIIMVYVGLISNICFMRGVAILTYMLLGLGFLRR
  VAPDFDLAGEEQLGKPRPTSCLCFMRHLNTHAESIAFGGAREKAVSFLLIA
  LAIAGFWWVIFESKLEVNLELFYHLCDDGLFMSNVASLMWDKPKRALLDLMLLR
  KKLWYGLDDPVTQKLPNNVTGSLIYALEHKGDRALYSTOVKYLHVLWMLLPGE
  LAHALRVLSVSSQSFVAFGDILZLHKHFLLESGGINRIFELDEFIDASQSGTSENQ
  TSLRDSQDLISFSEVDIITPAQKLMAKSLCEIVSGSLVLTGFGMSGKTSVFRVLRD
  IWTVCVCLTKPSLDIKELSGNGMPFVQRPYTCGLTQDIIYFLSKEEAKRAAK
  LYTSGESSTAGSILDSHLKXLTILENVELVYLLERDVGWDATTNWDIILSLGEQQLRG
  MVSLL"

```

```

exon      16239..16300
           /gene="AT4G39850"
           /number=1
intron    16301..16393
           /gene="AT4G39850"
           /number=1
exon      16394..16679
           /gene="AT4G39850"
           /number=2
intron    16680..16980
           /gene="AT4G39850"
           /number=2
exon      16981..17205
           /gene="AT4G39850"
           /number=3
intron    17206..17340
           /gene="AT4G39850"
           /number=3
exon      17341..17538
           /gene="AT4G39850"
           /number=4
intron    17539..17674
           /gene="AT4G39850"
           /number=4
exon      17675..18129
           /gene="AT4G39850"
           /number=5

Query Match      98.5%; Score 411.8; DB 8; Length 112067;
Best Local Similarity 99.5%; Pred. NO. 2.4e-80;
Matches 413; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  TGAGGTACGATCTTCGATCCCTTTGATTTTCCTCGAATATTTTTCGGTGATCGTGA 60
DB      85668 TGAGGTACGATCTTCGATCCCTTTGATTTTCCTCGAATATTTTTCGGTGATCGTGA 85609

QY      61  AACTACTGGAATCGCTCGATAGTGTGATCGAATATAGGCGAGATAGTTCTTATCTTGG 120
DB      85608 AACTACTGGAATCGCTCGATAGTGTGATCGAATATAGGCGAGATAGTTCTTATCTTGG 85549

QY      121  CCAATATCTTTGTTTTCGCGGATGATCTTCGATATAGGATTTTGTAGGTAGAGATGAA 180
DB      85548 CCAATATCTTTGTTTTCGCGGATGATCTTCGATATAGGATTTTGTAGGTAGAGATGAA 85489

QY      181  TCGTATAGCTAGATTTTCATCACCAGATAGTTCTTGTCTAGAAATCTCGAATCTCGA 240
DB      85488 TCGTATAGCTAGATTTTCATCACCAGATAGTTCTTGTCTAGAAATCTCGAATCTCGA 85429

QY      241  TAGTTTTCATGTTGAAATAGATTTTCTTATTCGCGATTTCTTGTAGGCTTTTGTAT 300
DB      85428 TAGTTTTCATGTTGAAATAGATTTTCTTATTCGCGATTTCTTGTAGGCTTTTGTAT 85369

QY      301  TTTCTTGATTTGATTTGCAATTAGGATTTTCTTGTGTTTGTGTTGATCTTACGATA 360
DB      85368 TTTCTTGATTTGATTTGCAATTAGGATTTTCTTGTGTTTGTGTTGATCTTACGATA 85309

QY      361  CATTCTCGCAATGCAATACGATGATGATCTAAATCTGTTAAATTTGTTGACAGAT 415
DB      85308 CATTCTCGCAATGCAATACGATGATGATCTAAATCTGTTAAATTTGTTGACAGAT 85254

RESULT 5
ATT5J17/c 118267 bp DNA linear PLN 21-SEP-1999
LOCUS Arabidopsis thaliana DNA chromosome 4, BAC clone T5J17 (ESSA
DEFINITION project).
ACCESSION AL035708
VERSION AL035708.2 GI:5918309
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

```

REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi  
1  
Bevan, M., Murphy, G., Ridley, P., Hudson, S., Bancroft, I., Mewes, H.W.,  
Mayer, K.F.X., Lemcke, K. and Schueller, C.  
Unpublished  
2 (bases 1 to 118267)  
EU Arabidopsis sequencing, project.  
Direct Submission  
Submitted (21-SEP-1999) MIPS, at the Max-Planck-Institut fuer  
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:  
schullemips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project  
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge  
Laboratory, John Innes Centre, Colney Lane, NR4 7UU Norwich, UK,  
E-mail: michael.bevan@bbsrc.ac.uk

## COMMENT

On Sep 22, 1999 this sequence version replaced gi:4490734.  
Information on performance of analysis and a more detailed  
annotation of this entry and other sequences of chromosomes 3, 4  
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

## FEATURES

Location/Qualifiers

## Source

1. 118267  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"

## misc\_feature

1. 22789  
/note="position 1-22789 overlaps to BAC T19P19, EMBL  
accession AL022605; please refer to this entry for  
analysis and annotation"

## gene

22854..24209

## CDS

22854..24209  
/gene="T5J17.10"  
/note="T5J17.10"  
/note="contains EST gb:F13826, Z46537"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB38897.1"  
/db\_xref="GI:4490735"  
/db\_xref="SPTREMBL:Q9SMR9"  
/translation="MSEPKLCFRKVLILLIYSLFLFSSSSPSSVDDHELORIDP  
HFRVRLVLDLSDTDDDDPPPKKKLLDVSSSSSSGTGKNTKLLKPLPS  
SSSTKQTKLAKTTGTGTHLNSTKSSNTTKTSSELKNSGTSTNSTSIKKS  
ADLSKSSSKNKTIIKPPSSKSSPSSKSPKSKVTSKQSEKIKFPLDDEED  
EDFVSERDLRFRSLIPDLERISTTSKNYINKANKQITKQFKYKNTAPTAS  
VVSVEVLPLLAVLLIFNRFKAYFLQKILFIQIVLSIVFLCLSSLVGTLEPK  
VLSVSSSTVYCLQTLQTLGYFVLLLLMLVAVFSTDCGLGKVLGLAQTGVGFAK  
GLHYVAVFHRVLRQPPKTNKWLHGVTATCFLILCLLSNERRKKYLEBGGDEGKK  
N"

## exon

22854..24209  
/gene="T5J17.10"  
/number=1

## misc\_feature

26106..26772  
/note="EST N95877 matches to positions 26106-26355,  
26723-26772; possibly related to subsequent gen"

## gene

26806..33879

## CDS

26806..33879  
/gene="T5J17.20"  
/note="26806..26867,26961..27246,27548..27772,27908..28105,  
28242..28696,28789..28924,29000..29078,29606..29838,  
29893..30197,30279..30411,30468..30623,30780..31002,  
31082..31190,31412..31502,31772..31939,32020..32135,  
32193..32427,32491..32671,32729..32929,33101..33247,  
33317..33550,33701..33879"  
/gene="T5J17.20"  
/note="similarity to 70K peroxisomal membrane protein,  
Rattus norvegicus, PIR2.A35723  
Contains ABC transporters family signature  
[LSLGEQRLGNARLF][LSLGEQRLGNVSLJ]"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB38898.1"  
/db\_xref="GI:4490736"  
/db\_xref="GOA:Q9SMR8"  
/db\_xref="SPTREMBL:Q9SMR8"



QY 361 CATCTCTGCAATTGAATACGATGATGATCAATTAATCTTTGTTAAATTTCTTGAACAGAT 415  
DB 95875 CATTCGGCAATGATGATGATGATCAATTAATCTTTGTTAAATTTCTTGAACAGAT 95821

RESULT 6  
AX346536  
LOCUS AX346536 5998 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 1607 from Patent WO0200928.  
ACCESSION AX346536  
VERSION AX346536.1 GI:18494422  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.

REFERENCE 1  
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of diseases associated with the immune system  
JOURNAL Patent: WO 0200928-A 1607 03-JAN-2002;  
EpiGenomics AG (DE)  
FEATURES Location/Qualifiers  
source 1..5998  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN  
Query Match 13.0%; Score 54.4; DB 6; Length 5998;  
Best Local Similarity 49.7%; Pred. No. 0.045;  
Matches 165; Conservative 0; Mismatches 166; Indels 1; Gaps 1;  
QY 78 GATAGTGTGACGAATAATAGGCGAGATTAGTTCTTATCTTGGCCATTAATCTTTCTT 137  
DB 4382 GAGTAGTGGTTTGGTATATAGTTAAATTTGATTTTCGAATTTTGGGTTTAAATGATTTT 4441  
QY 138 CGCGGATGATCTCCGTA- TAAAGATTTTAGTTAGAGATGAATCGTATAGTAGATT 196  
DB 4442 TGTTTATGTTTGGATTTAGTTGGGATATAGTTATAGTTATGTTTGTAGTTAATTA 4501  
QY 197 CATCACAGATAGTTCTTTTGTCTAGAAATCTCTGAATTTCTCGATAGTTTTCACATGTT 256  
DB 4502 AAAAAAATTTGTTTATTTTATAGAGATGAAGTTTATGTTTGTAGTTGG 4561  
QY 257 AATAGATTTCTTATTCGGCATTTGTTGATTTAGTTGTTTCTTGTATGCGA 316  
DB 4562 TTTTGAATTTTGGTTTAGCGGATTTTATTTTATTTTGTAGTTTGTGTTATATA 4621  
QY 317 TTGCAATTAGGGATTTTCTTTGTTTGTGTTGATCTTACGATACATTCCTGCAATGAA 376  
DB 4622 TTTTATGGGPAATTTTATTTGTTATTTATTTATTTTATTTTATTTTATTTT 4681  
QY 377 TAGGTATGATCTAAATCTTTGTTAAATTTGTTG 408  
DB 4682 TATGGGATTTATTTTGTGTTTATGTTTATG 4713

RESULT 7  
AX344561  
LOCUS AX344561 349980 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 12 from Patent WO0200932.  
ACCESSION AX344561  
VERSION AX344561.1 GI:18492447  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.

REFERENCE 1  
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of known genetic parameters within the mhc  
JOURNAL Patent: WO 0200932-A 12 03-JAN-2002;  
EpiGenomics AG (DE)  
FEATURES Location/Qualifiers

source 1..349980  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"

Query Match 12.4%; Score 52; DB 6; Length 349980;  
Best Local Similarity 47.8%; Pred. No. 0.091;  
Matches 151; Conservative 0; Mismatches 165; Indels 0; Gaps 0;  
QY 38 AATATTTTTCGGTATCGTGAACACTACTGGAATCTCGTAGTGGTACGAATAG 97  
DB 277723 AATATTTTTCGGTATCGTGAACACTACTGGAATCTCGTAGTGGTACGAATTT 277782  
QY 98 GCGAGATTGTTCTATCTTGGCCATTTCTTTTCGCGAATGATCTTCCGTAT 157  
DB 277783 TTTAGTAAATTTTATTTTTCGATGTTTATTTTATTTTTCGAGTTTAAATAT 277842  
QY 158 AAGATTTTAGTTAGAGATGAATCGTATAGTATTTTATCATCACCAGATTTCTTG 217  
DB 277843 ATATATTATGAAATAATTTGGAGTTTTTATTTTATTTATATAGATAAGTTTATTA 277902  
QY 218 TCTAGATCTCTGAATCTCGATAGTTTTCACATGTTGAATAGTTTCTTATTCGG 277  
DB 277903 TATTGGTAAATTTAAATATATATTAGTATTTATTTTATTTTATTTTATTTT 277962  
QY 278 CGATTGTTGATTAGGTTTTCGATTTTCTTTCGATTCGATTCGATTCGATTTCTTT 337  
DB 277963 TTTATTTATTTTGGTTATTTTATTTTTCGTTCTAGATTTTATTTTATTTTAT 278022  
QY 338 GGTGTTGTTGATCT 353  
DB 278023 TTGTTGTTTATATTT 278038

RESULT 8  
AX014842  
LOCUS AX014842 253151 bp DNA linear INV 07-OCT-2002  
DEFINITION Plasmodium falciparum 3D7 chromosome 11 section 7 of 8 of the complete sequence.  
ACCESSION AX014842  
VERSION AX014842.1 GI:23496321  
KEYWORDS Plasmodium falciparum 3D7  
SOURCE Plasmodium falciparum 3D7  
ORGANISM Plasmodium falciparum 3D7  
REFERENCE 1 (bases 1 to 253151)  
AUTHORS Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., James, K., Eissen, J.A., Paulsen, I.T., Salzberg, S.L., Craig, A., Carlton, J.M., Paine, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., Shallen, S.J., Suh, B., Peterson, J., Anguoli, S., Perce, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M.A., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M. and

|               |   |
|---------------|---|
| TITLE         | Barrell,B.  |
| JOURNAL       | Genome sequence of the human malaria parasite Plasmodium falciparum   |
| REFERENCE     | Nature 419 (6906), 498-511 (2002)   |
| AUTHORS       | 12368864  |
| JOURNAL       | 2 (bases 1 to 253151)   |
| FEATURES      | Direct Submission   |
| source        | Submitted (13-SEP-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA |
| repeat_region | Location/Qualifiers   |
| repeat_region | 1. .253151  |
| repeat_region | /organism="Plasmodium falciparum 3D7"   |
| repeat_region | /mol_type="genomic DNA"   |
| repeat_region | /isolate="3D7"  |
| repeat_region | /db_xref="taxon:36329"  |
| repeat_region | /chromosome="11"  |
| repeat_region | complement(73..101)   |
| repeat_region | /note="AT rich"   |
| repeat_region | /rpt_type=tandem  |
| repeat_region | 173..279  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | /rpt_unit="(TAAAAA)n"   |
| repeat_region | complement(302..328)  |
| repeat_region | /note="AT rich"   |
| repeat_region | /rpt_type=tandem  |
| repeat_region | complement(393..466)  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | /rpt_unit="(TA)n"   |
| repeat_region | 393..442  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | 440..466  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | 545..578  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | complement(545..578)  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | /rpt_unit="(TA)n"   |
| repeat_region | complement(580..673)  |
| repeat_region | /note="AT rich"   |
| repeat_region | /rpt_type=tandem  |
| repeat_region | 762..795  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | complement(762..795)  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | /rpt_unit="(TA)n"   |
| repeat_region | complement(869..955)  |
| repeat_region | /note="AT rich"   |
| repeat_region | /rpt_type=tandem  |
| repeat_region | complement(956..975)  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | /rpt_unit="(A)n"  |
| repeat_region | complement(983..1040)   |
| repeat_region | /note="AT rich"   |
| repeat_region | /rpt_type=tandem  |
| repeat_region | 1094..1121  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | /rpt_unit="(A)n"  |
| repeat_region | complement(1149..1221)  |
| repeat_region | /note="AT rich"   |
| repeat_region | /rpt_type=tandem  |
| repeat_region | 1239..1281  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | complement(1243..1319)  |
| repeat_region | /rpt_type=tandem  |
| repeat_region | /rpt_unit="(TA)n"   |
| repeat_region | complement(1323..1379)  |
| repeat_region | /note="AT rich"   |
| repeat_region | /rpt_type=tandem  |
| repeat_region | complement(1571..1605)  |
| repeat_region | /note="AT rich"   |
| repeat_region | /rpt_type=tandem  |
| repeat_region | complement(1621..1699)  |
| repeat_region | /note="AT rich"   |

```
repeat_region 3845..3884
/rpt_type=tandem
repeat_region 3897..3944
/rpt_type=tandem
/rpt_unit="(TATAA)n"
complement(3946..3999)
/note="AT-rich"
/rpt_type=tandem
4000..4024
/rpt_type=tandem
/rpt_type=tandem
complement(4000..4024)
/rpt_type=tandem
/rpt_unit="(A)n"
complement(4026..4104)
/note="AT-rich"
/rpt_type=tandem
4163..4226
/rpt_type=tandem
/rpt_unit="(TAAAA)n"
4232..4267
/rpt_type=tandem
complement(4232..4267)
/rpt_type=tandem
/rpt_unit="(TA)n"
complement(4339..4359)
/rpt_type=tandem
/rpt_unit="(TA)n"
4380..4407
/rpt_type=tandem
complement(4380..4407)
/rpt_type=tandem
/rpt_unit="(A)n"
complement(4408..4457)
/note="AT-rich"
/rpt_type=tandem

Query Match 12.2%; Score 51.2; DB 3; Length 253151;
Best Local Similarity 49.6%; Pred. No. 0.14;
Matches 131; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 155 TATAAGATTTTAGGTAGATGATCGTATAGTAGATTCATCACCAGATAGTTCT 214
Db TATTATATATTTTATATATATACATACATATATATATATATATATATATATAT 3345

QY 215 TTGCTAGAACTCTGAAATCTCGATAGTTTTCATATGTTTCAATGTAAGATGTTCTATT 274
Db ATATAATTAATTTATATATATAAATAATATTTTATTTATTTATTTATTTATTTATTTATTT 3405

QY 275 CGCGATGTTGATAGGTTTGGATTTCTTCTGATTCGATTCGATTCGATTCGATTCGATTC 334
Db TGATATATTTTGTATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3465

QY 335 TTGCTTTTGTGTTGATCTTACGATACATTCCTGCAATTCGATTCGATTCGATTCGATTCGATTC 394
Db TATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3525

QY 395 TTGTTAATTTGTTGACAGATCCC 418
Db TTAATAATATCATGAATATTTCAC 3549

RESULT 9
LOCUS AX339160 17527 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 27 from Patent WO0176451.
ACCESSION AX339160
VERSION AX339160.1 GI:18129237
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial sequences
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1405 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES
source
1..17527
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 12.1%; Score 50.6; DB 6; Length 17527;
Best Local Similarity 52.1%; Pred. No. 0.27;
Matches 113; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 192 GATTTCAATCACCAGATAGTTCTTCTTCTAGAACTCTGAAATCTCGATAGTTTTCACA 251
Db GTTGTATAAGAAATATTTGTTTTTATTTAGTCTGTTATTTTATTTAGAAAGTTTAAAGA 16649

QY 252 TGTCTAAATAGATTTCTTATTTCGGCGATTTGATTAGGTTTTCGATTTCTTGATTA 311
Db GGTGTTATTTGTTGTTGTTTGTGTTGTTTATTTTATTTTATTTTATTTTATTTTATTTT 16709

QY 312 TCGAATTCGAATAGGATTTCTTCTGTTTGTGTTGATCTTACGATACATTCCTGCAA 371
Db TAATAGTGAATATTGGGAAGTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 16769

QY 372 TTGAATAGTATGATCTAAATCTTTAAATTTGTTG 408
Db TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 16806

RESULT 10
LOCUS AX346334 17527 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 1405 from Patent WO0200928.
ACCESSION AX346334
VERSION AX346334.1 GI:18494220
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial sequences
REFERENCE 1
AUTHORS Olek,A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1405 03-JAN-2002;
EpiGenomics AG (DE)
FEATURES
source
1..17527
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 12.1%; Score 50.6; DB 6; Length 17527;
Best Local Similarity 52.1%; Pred. No. 0.27;
Matches 113; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 192 GATTTCAATCACCAGATAGTTCTTCTTCTAGAACTCTGAAATCTCGATAGTTTTCACA 251
Db GTTGTATAAGAAATATTTGTTTTTATTTAGTCTGTTATTTTATTTAGAAAGTTTAAAGA 16649

QY 252 TGTCTAAATAGATTTCTTATTTCGGCGATTTGATTAGGTTTTCGATTTCTTGATTA 311
Db GGTGTTATTTGTTGTTGTTTGTGTTGTTTATTTTATTTTATTTTATTTTATTTTATTTT 16709

QY 312 TCGAATTCGAATAGGATTTCTTCTGTTTGTGTTGATCTTACGATACATTCCTGCAA 371
Db TAATAGTGAATATTGGGAAGTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 16769

QY 372 TTGAATAGTATGATCTAAATCTTTAAATTTGTTG 408
Db TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 16806
```



Db 16770 TTTT...TTTATAG 16806

RESULT 11  
AX278060  
LOCUS  
DEFINITION  
Sequence 223 from Patent WO0177375.  
ACCESSION  
AX278060  
VERSION  
AX278060.1 GI:16605132  
KEYWORDS  
synthetic construct  
synthetic construct  
ORGANISM  
artificial sequences.  
REFERENCE  
1  
Olek, A., Piepenbrock, C. and Berlin, K.  
Diagnosis of diseases associated with gene regulation  
Patent: WO 0177375-A 223 18-OCT-2001;  
Epigenomics AG (DE)  
JOURNAL  
Location/Qualifiers  
FEATURES  
1. .6306  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN  
Query Match 12.0%; Score 50.2; DB 6; Length 6306;  
Best Local Similarity 50.2%; Pred. No. 0.37;  
Matches 124; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
QY 161 GATTTAGTTAGATGAATCGTATAGTATTCATCACCAGATAGTTCTTGTCT 220  
DB 5490 GAGTTTCGGTTCGAGGGGTATAGTATGTTAGTATTAATTTGTTTAGGTAG 5549  
QY 221 AGAATCTCTGAAATCTCGATAGTTTTCACATGCTGTAATAGATTTGTTTATCGGCGA 280  
DB 5550 ATGTTTCGGTATGATTTTGAATTTTGTATTTTATGTTTATTTTAAAGTT 5609  
QY 281 TTGTTGATTAGGGTTTTCGATTTCTTGATTATGCAATTCGCAATAGGATTTCTTGGT 340  
DB 5610 TGAATTAATTTATTTTGGTGGGTTTATATATCGTTTGTATTTTGTATTTT 5669  
QY 341 TTGTTGTTGATCTTACGATACATTCCTGCAATGATGATGATCTAAATCTTGTA 400  
DB 5670 TTTT...TTTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTT 5729  
QY 401 ATTGTT 407  
DB 5730 TTTT...TTT 5736

RESULT 12  
AX323843  
LOCUS  
DEFINITION  
Sequence 331 from Patent WO0192565.  
ACCESSION  
AX323843  
VERSION  
AX323843.1 GI:18094592  
KEYWORDS  
synthetic construct  
synthetic construct  
artificial sequences.  
ORGANISM  
artificial sequences.  
REFERENCE  
1  
Olek, A., Piepenbrock, C. and Berlin, K.  
Diagnosis of diseases associated with dna transcription  
Patent: WO 0192565-A 331 06-DEC-2001;  
Epigenomics AG (DE)  
JOURNAL  
Location/Qualifiers  
FEATURES  
1. .6306  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="Chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 12.0%; Score 50.2; DB 6; Length 6306;  
Best Local Similarity 50.2%; Pred. No. 0.37;  
Matches 124; Conservative 0; Mismatches 123; Indels 0; Gaps 0;  
QY 161 GATTTAGTTAGATGAATCGTATAGTATTCATCACCAGATAGTTCTTGTCT 220  
DB 5490 GAGTTTCGGTTCGAGGGGTATAGTATGTTAGTATTAATTTGTTTAGGTAG 5549  
QY 221 AGAATCTCTGAAATCTCGATAGTTTTCACATGCTGTAATAGATTTGTTTATCGGCGA 280  
DB 5550 ATGTTTCGGTATGATTTTGAATTTTGTATTTTATGTTTATTTTAAAGTT 5609  
QY 281 TTGTTGATTAGGGTTTTCGATTTCTTGATTATGCAATTCGCAATAGGATTTCTTGGT 340  
DB 5610 TGAATTAATTTATTTTGGTGGGTTTATATATCGTTTGTATTTTGTATTTT 5669  
QY 341 TTGTTGTTGATCTTACGATACATTCCTGCAATGATGATGATCTAAATCTTGTA 400  
DB 5670 TTTT...TTTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTT 5729  
QY 401 ATTGTT 407  
DB 5730 TTTT...TTT 5736

RESULT 13  
BX470257/c  
LOCUS  
DEFINITION  
Danio rerio clone DKEY-99015, WORKING DRAFT SEQUENCE, 5 unordered pieces.  
ACCESSION  
BX470257  
VERSION  
BX470257.3 GI:35209325  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE  
Danio rerio (zebrafish)  
ORGANISM  
Danio rerio  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
REFERENCE  
1 (bases 1 to 208653)  
McLaren, S.  
Direct Submission  
Submitted (24-SEP-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Sep 24, 2003 this sequence version replaced gi:32169046.  
COMMENT  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: zfish-help@sanger.ac.uk  
----- Project Information  
Center project name: zK99015  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Chemistry: Dye-terminator; 100% of reads  
Consensus quality: 205905 bases at least Q40  
Consensus quality: 206063 bases at least Q30  
Consensus quality: 206151 bases at least Q20  
Insert size: 206253; sum-of-contigs  
Insert size: 198710; 1.2% error; agarose-fp  
Quality coverage: 7.80x in Q20 bases; sum-of-contigs Quality  
coverage: 8.14x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 64965; contig of 64965 bp in length

\* 64966 65065: gap of 100 bp  
 \* 65066 83771: contig of 15706 bp in length  
 \* 83772 83771: gap of 100 bp  
 \* 83872 136962: contig of 53091 bp in length  
 \* 136963 137062: gap of 100 bp  
 \* 137063 186755: contig of 49693 bp in length  
 \* 186756 186855: gap of 100 bp  
 \* 186856 206653: contig of 13798 bp in length.

## FEATURES

location/Qualifiers  
 1..206653  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone\_PDXKEY="99015"  
 /clone\_lib="DanioKey"  
 1..64965  
 /note="assembly\_fragment:03254  
 fragment\_chain:1"  
 65066..83771  
 /note="assembly\_fragment:01800  
 fragment\_chain:1"  
 83872..136962  
 /note="assembly\_fragment:02854  
 fragment\_chain:1"  
 137063..186755  
 /note="assembly\_fragment:00600.0"  
 186856..206653  
 /note="assembly\_fragment:01616.0"

## ORIGIN

Query Match 11.9%; Score 49.6; DB 2; Length 206653;  
 Best Local Similarity 47.7%; Pred. No. 0.33;  
 Matches 145; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 104 TTAGTTCTATCTTGGCCATATCTGTCTTCGCCGATGATCTCCGATATAAGAT 163  
 Db TTTTATTTTATTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 33739  
 QY 164 TTTAGTTAGATGATCGTATAGCTATGATTCATCCAGATAGTTCTTGTCTAGA 223  
 Db ATTATTTATTTTATTTTATTTGCTATTTATTTTATTTTATTTTATTTTATTTATTA 33679  
 QY 224 ATCTCTGAATCTCGATAGTTTTCACATGCTGAATAGTTCTTATTCGGCGATTG 283  
 Db TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 33619  
 QY 284 TTGATTTAGGTTTGTGATTTCTTGATTCGATTCGATTCGATTCGATTCGATTTT 343  
 Db TTTTTCATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 33559  
 QY 344 GTGTTGATCTAGTACATCTCGAATGATGATGATGATGATGATGATGATGATGAT 403  
 Db GTTTTGTATTTGATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 33499  
 QY 404 TGT 407  
 Db 33498 AGTT 33495

## RESULT 14

AC141780 127172 bp DNA linear HTG 19-MAR-2003  
 LOCUS Apis mellifera clone CH224-59E13, \*\*\* SEQUENCING IN PROGRESS \*\*\*  
 DEFINITION 51 unordered pieces.  
 ACCESSION AC141780  
 VERSION AC141780.1 GI:29123964  
 KEYWORDS HTG; HTGS PHASE1.  
 SOURCE Apis mellifera (honeybee)  
 ORGANISM Apis mellifera  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;  
 Apidae; Apis.  
 REFERENCE 1 (bases 1 to 127172)

## AUTHORS

Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
 Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,  
 Barbara, J., Benton, J., Binage, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Benn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gortell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Haylak, P., Haves, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
 Jacobs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louiseged, H.,  
 Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokoko, S., Ogih, M., Okwuonu, G.,  
 Ozagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherren, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,  
 Sodergren, E., Sonalke, I., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinsteck, G. and Gibbs, R.

## TITLE

JOURNAL

## REFERENCE

2 (bases 1 to 127172)

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
 Project Information  
 Center project name: AMCV  
 Center clone name: CH224-59E13  
 Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 109239 bases at least Q40  
 Consensus quality: 114958 bases at least Q30  
 Consensus quality: 118059 bases at least Q20  
 Estimated insert size: 102169; sum-of-contigs estimation  
 Quality coverage: 2x in Q20 bases; sum-of-contigs estimation  
 \*\*\*\*\*  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 51 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.



## artificial sequences.

## REFERENCE

1  
Olek, A., Piepenbrock, C. and Berlin, K.  
Diagnosis of known genetic parameters within the mhc  
Patent: WO 0200932-A 24 03-JAN-2002;  
Epigenomics AG (DE)

## FEATURES

## source

1. 349980  
Location/Qualifiers  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"

/db\_xref="taxon:32830"  
/notes="chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3.673778 <223>-split as follows--seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <223>-Original length of seq 2: 3.673778 <223>-split as follows--seq 14 0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16 600.001 949.980-seq 17 900.001 1.249.980-seq 18 1.200.001 1.549.980-seq 19 1.500.001 1.849.980-seq 20 1.800.001 2.149.980-seq 21 2.100.001 2.449.980-seq 22 2.400.001 2.749.980-seq 23 2.700.001 3.049.980-seq 24 3.000.001 3.349.980-seq 25 3.300.001 3.649.980-seq 26 3.600.001 3.673.778"

## ORIGIN

Query Match 11.8%; Score 49.2; DB 6; Length 349980;  
Best Local Similarity 45.2%; Fred. No. 0.37; Indels 0; Gaps 0;  
Matches 180; Conservative 0; Mismatches 218;  
QY 10 ATTCTTCGATCCCTTTGATTTTCCTCGAAATATTTTTCGGTGAATCGTGAACACTACTCG 69  
Db 317812 AATTTTGTGTTTTCGTTTTTTTTTTTAAATTAATTTTGGTAGTATTAATATGAA 317871  
QY 70 AATCGCTCGATAGTGTAGCAAAATAGGCGAGATAGTTCTTATCTTGGCCATTATCT 129  
Db 317872 AATTAATTTTGTAGTTTTAGTATTAATTTTTCGATTTTTTTTTTTTTTAGATTTTT 317931  
QY 130 TGTTCCTTCGCGAATGATCTTCGTATAAAGATTTTAGGTAGAGATGAATCGTATAGC 189  
Db 317932 ATTTTTTTTTTTTTTAAATAAATAAATAATTTTTTTTATTTTAAAGTTTAGTTTTT 317991  
QY 190 TAGATTTTCATCCAGATAGTTCTTTGTCTAGAAATCTCGAAATCTCGATAGTTTCA 249  
Db 317992 TATTTTTTATTTATTTATTTTTCGTAATTTTGGTATTTTGGTATTTTAGTAGTAAGTATTATTAAT 318051  
QY 250 CATGTGTAATAGATTTCTTCTTATTCGCGATTTCTGATTAGGTTTGGATTTCTTGAT 309  
Db 318052 TATTAATTAATTAATTTTTTTTAAATTTTAAAGATTAAGTTTAATGTAATTTTTTTTTT 318111  
QY 310 TATCGATTCGAATAGGATTTCTTTGGTTTGTGTTGATCTTAGACATATCTCTGC 369  
Db 318112 TGAAGTAATTAATATATGTTTTCGATAAATAATTTTTTGTATTTTATTTATATTTTGT 318171  
QY 370 AATTGATACGTATGGATCTAAATCTGTTAATTTGT 407  
Db 318172 TTTTGAATTTTAAATAATTTTTTTTTTATTTTTTTTTT 318209

Search completed: June 20, 2004, 03:45:45  
Job time : 2934.58 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 23:55:13 ; Search time 309.833 Seconds

(without alignments)

5731.297 Million cell updates/sec

Title: US-09-000-062-6

Perfect score: 418

Sequence: 1 tggagtagattcttcgac.....taattgtgacagatccc 418

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 337863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 29Jan04:\*

1: Geneseqn1980s:\*

2: Geneseqn1990s:\*

3: Geneseqn2000s:\*

4: Geneseqn2001bs:\*

5: Geneseqn2001bs:\*

6: Geneseqn2002s:\*

7: Geneseqn2003as:\*

8: Geneseqn2003bs:\*

9: Geneseqn2003cs:\*

10: Geneseqn2004s:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 418   | 100.0       | 418    | 2  | Aat85996 Arabidops |
| 2          | 54.4  | 13.0        | 5998   | 6  | ABL33634 Human inm |
| 3          | 50.6  | 12.1        | 17527  | 6  | ABL33432 Human inm |
| 4          | 50.6  | 12.1        | 17527  | 6  | AAS63332 Chemicall |
| 5          | 50.2  | 12.0        | 6306   | 4  | AAS45515 Chemicall |
| 6          | 50.2  | 12.0        | 6306   | 4  | ABK28457 DNA trans |
| 7          | 49    | 11.7        | 5987   | 6  | ABL33562 Human inm |
| 8          | 49    | 11.7        | 5987   | 6  | ABQ67101 Human ang |
| 9          | 49    | 11.7        | 5987   | 9  | ADB54307 Pretreate |
| 10         | 49    | 11.7        | 5987   | 9  | ADB54179 Pretreate |
| 11         | 48.8  | 11.7        | 15479  | 6  | ABK39964 Human che |
| 12         | 48.4  | 11.6        | 17848  | 4  | AAS45323 Chemicall |
| 13         | 48.4  | 11.6        | 17848  | 6  | ABK39976 Human che |
| 14         | 48.4  | 11.6        | 17848  | 6  | ABK28164 DNA trans |
| 15         | 48    | 11.5        | 2501   | 7  | ABZ10074 Haematopo |
| 16         | 48    | 11.5        | 2501   | 7  | ABZ10220 Haematopo |
| 17         | 48    | 11.5        | 6059   | 4  | ABL33480 Human inm |
| 18         | 48    | 11.5        | 8245   | 4  | AAS46448 Tumour su |
| 19         | 48    | 11.5        | 16724  | 6  | ABL33090 Human inm |
| 20         | 48    | 11.5        | 16724  | 6  | ABL34536 Human met |
| 21         | 48    | 11.5        | 16724  | 6  | ABL70259 Chemicall |
| 22         | 47.6  | 11.4        | 5649   | 9  | ADB54254 Pretreate |
| 23         | 47.6  | 11.4        | 5649   | 9  | ADE84184 Human lym |

|    |      |      |       |   |                     |
|----|------|------|-------|---|---------------------|
| 24 | 47.4 | 11.3 | 21537 | 6 | ABL33399            |
| 25 | 47.2 | 11.3 | 5371  | 6 | ABL34295            |
| 26 | 46.8 | 11.2 | 714   | 6 | ABQ18144            |
| 27 | 46.8 | 11.2 | 714   | 6 | ABQ18145            |
| 28 | 46.8 | 11.2 | 8805  | 6 | ABK40016            |
| 29 | 46.6 | 11.1 | 6012  | 6 | ABL34564 Human met  |
| 30 | 46.6 | 11.1 | 6012  | 6 | ABL34564            |
| 31 | 46.6 | 11.1 | 6668  | 6 | ABL33697 Human inm  |
| 32 | 46.2 | 11.1 | 6048  | 6 | ABL332509           |
| 33 | 46.2 | 11.1 | 18817 | 6 | ABL34495 Human met  |
| 34 | 46.2 | 11.1 | 18817 | 6 | ABL70162 Chemicall  |
| 35 | 46.2 | 11.1 | 26811 | 2 | AAX20253 Borrelia   |
| 36 | 46   | 11.0 | 5649  | 4 | AAS46384 Tumour su  |
| 37 | 46   | 11.0 | 5649  | 6 | ABK40008 Human che  |
| 38 | 46   | 11.0 | 5649  | 6 | ABL32849            |
| 39 | 46   | 11.0 | 5649  | 9 | ADB54126 Pretreate  |
| 40 | 46   | 11.0 | 5649  | 9 | ADB54108 Human lym  |
| 41 | 46   | 11.0 | 6161  | 6 | ABL32623 Human inm  |
| 42 | 46   | 11.0 | 7892  | 6 | ABK40056 Human che  |
| 43 | 45.8 | 11.0 | 6223  | 6 | AAS61176 Human gen  |
| 44 | 45.6 | 10.9 | 374   | 7 | ABX47736 Bovine ES  |
| 45 | 45.6 | 10.9 | 593   | 6 | ABQ20161 Oligonuecl |

ALIGNMENTS

RESULT 1

AAT85996

ID AAT85996 standard; DNA; 418 BP.

XX:

AC AAT85996;

XX

DT 17-NOV-1997 (first entry)

XX

DE Arabidopsis thaliana histone H3.3-like DNA fragment (intron 1).

XX

Plant expression regulation sequence; intron 1; histone;

KW herbicide tolerance; 5-enolpyruvylshikimate-3-phosphate synthase; EPSPS;

KW glyphosate; ds.

XX

OS Arabidopsis thaliana.

XX

Key Location/Qualifiers

intron

FT

FT

FT

XX

PN WO9704114-A2.

XX

PD 06-FEB-1997.

XX

PF 17-JUL-1996; 96WO-FR001109.

XX

PR 19-JUL-1995; 95FR-00008980.

XX

PA (RHON ) RHONE POULENC AGROCHIMIE.

XX

PI Derose R, Chaubet N, Gigot C;

XX

DR WPI; 1997-132652/12.

XX

PT New regulatory sequence for chimeric gene expression in rapidly growing parts of a plant - includes at least one intron from a plant histone gene and is useful for imparting resistance to herbicides.

XX

PS Claim 4; Page 26; 31pp; French.

XX

CC The known cosmid clone c22 of Arabidopsis thaliana contains two histone H3.3-like genes. Digestion of clone c22 with restriction enzyme DdeI, Klenow fragment and MspI generated a fragment of 418 bp having the present sequence. This fragment, designated intron 1, was ligated to synthetic linkers for cloning into plant expression vectors. In addition

CC to the intron 1 sequence, the vectors contained a plant promoter and a  
CC herbicide tolerance gene (e.g. a mutated version of the 5'  
CC enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene). The intron  
CC enhances expression of the herbicide tolerance gene in rapidly growing  
CC parts of plants. The intron can also be used to enhance expression of  
CC genes that impart resistance to pathogens or that encode nutritional or  
CC therapeutic proteins  
XX Sequence 418 BP; 100 A; 63 C; 85 G; 170 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 418; DB 2; Length 418;  
Best Local Similarity 100.0%; Pred. No. 2.8e-92;  
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAGGTAGATCTTCGATCTCTTGGATTTCTCGTGAATATTTTCGTTGATCGTGA 60  
DB 1 TGAGGTAGATCTTCGATCTCTTGGATTTCTCGTGAATATTTTCGTTGATCGTGA 60  
QY 61 AACTACTCGAATCGCTCGATAGGTGTGTACGAAATAGCGAGATAGTTCTATCTTGG 120  
DB 61 AACTACTCGAATCGCTCGATAGGTGTGTACGAAATAGCGAGATAGTTCTATCTTGG 120  
QY 121 CCAATATCTTGTCTTTCGCGGATGATCTCCGTATAAAGATTTAGGTTAGAGATGAA 180  
DB 121 CCAATATCTTGTCTTTCGCGGATGATCTCCGTATAAAGATTTAGGTTAGAGATGAA 180  
QY 181 TCGTATAGCTAGATTTTCATCACCAGATAGTTCTTGTCTAGAAATCTCGAAATCTCGA 240  
DB 181 TCGTATAGCTAGATTTTCATCACCAGATAGTTCTTGTCTAGAAATCTCGAAATCTCGA 240  
QY 241 TAGTTTTCATGTTGTAATAGATGTTCTTATTCGGCGATGTTGATAGGTTTGGAT 300  
DB 241 TAGTTTTCATGTTGTAATAGATGTTCTTATTCGGCGATGTTGATAGGTTTGGAT 300  
QY 301 TTTCCTGATTATGCGATTCGATAGGATTTCTTGGTTTGGTTGATCTTACGATA 360  
DB 301 TTTCCTGATTATGCGATTCGATAGGATTTCTTGGTTTGGTTGATCTTACGATA 360  
QY 361 CATTCTCGAATGATACGATAGATCTAAATCTTGTAAATTTGTTGAAACAGATCCC 418  
DB 361 CATTCTCGAATGATAGATAGATCTAAATCTTGTAAATTTGTTGAAACAGATCCC 418  
RESULT 2  
ABL33634  
ID ABL33634 standard; DNA; 5998 BP.  
XX  
AC ABL33634;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 1607.  
XX  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosolic; nontropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200200928-A2.  
XX  
XX  
PD 03-JAN-2002.  
XX  
XX 02-JUL-2001; 2001WO-EP007537.  
XX  
XX 30-JUN-2000; 2000DB-01032529.  
PR 01-SEP-2000; 2000DB-01043826.  
XX

PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.  
XX  
XX Claim 1; SEQ ID NO 1607; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory ulcerative bowel  
CC diseases. The present sequence is a gene of the invention  
XX  
SQ Sequence 5998 BP; 1217 A; 92 C; 1806 G; 2883 T; 0 U; 0 Other;  
Query Match 13.0%; Score 54.4; DB 6; Length 5998;  
Best Local Similarity 49.7%; Pred. No. 0.0013;  
Matches 165; Conservative 0; Mismatches 166; Indels 1; Gaps 1;  
QY 78 GATGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 137  
DB 4382 GACTAGTGGTTGAGTATAGTAAATGTTGATGATGATGATGATGATGATGAT 4441  
QY 138 CGCGGATGATCTTCGTA-TAAGATTTTAGTTAGATGATGATGATGATGATGAT 196  
DB 4442 TGTTTAGTTTGTAGTGGGATGATGATGATGATGATGATGATGATGATGAT 4501  
QY 197 CATCACCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 256  
DB 4502 AAAAAAATGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 4561  
QY 257 AATAGATGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 316  
DB 4562 TTTTGAATTTTGGTTTGGGATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 4621  
QY 317 TTGCAATTAGGATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 376  
DB 4622 TTTTAGGTTAATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 4681  
QY 377 TACGTATGATCTAAATCTTGTAAATTTGTTG 408  
DB 4682 TATGGGATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 4713  
RESULT 3  
ABL33432  
ID ABL33432 standard; DNA; 17527 BP.  
XX  
XX ABL33432;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Human immune system associated gene SEQ ID NO: 1405.  
XX  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosolic; nontropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
ds.  
XX  
OS Homo sapiens.  
XX

```
PN WO200200928-A2.
XX
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 1405; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 17527 BP; 4501 A; 296 C; 4094 G; 8636 T; 0 U; 0 Other;
XX
XX Query Match 12.1%; Score 50.6; DB 6; Length 17527;
XX Best Local Similarity 52.1%; Pred. No. 0.013;
XX Matches 113; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
XX
XX QY 192 GATTTCATCACAGATAGTTCTTCTGCTAGAAATCTCTGAAATCTCGATAGTTTTCACA 251
XX Db 16590 GTTGTATAAGAAATATTTGTTTTTTTATTGAGTCGTATTATTAGAAAGTTTAAAGA 16649
XX
XX QY 252 TGTGTAATAGATGTTCTTATTGCGGATGTTGATTAGGGTTTGTATTCTTGATTA 311
XX Db 16650 GGTGGTATTGTTGGTGTGTTTTTTTGTGGTTTTTTTATTATTTTTTTTAGTTT 16709
XX
XX QY 312 TGGCATTCGAATAGGGAATTTCTTGGTTTGTGTGATCTTACGATACATCTCGCAA 371
XX Db 16710 TAATAGTGATTATTGGGAAGTTTTTTTTTGTGTTGTTTTTTTGTATTATTTTAT 16769
XX
XX QY 372 TTGAATACGATCGAATCAATCTGTTAATTTGTTG 408
XX Db 16770 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAG 16806
XX
XX
XX RESULT 4
XX AAS63332
XX ID AAS63332 standard; DNA; 17527 BP.
XX
XX AC AAS63332;
XX
XX 29-JAN-2002 (first entry)
XX
XX Chemically pretreated metabolism associated gene #27.
XX
XX Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;
XX solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;
XX single nucleotide polymorphism detection; SNP; stool; urine; lung;
XX cerebral-spinal fluid; intestine; brain; heart; prostate; breast; DUSP2;
XX EPHX2; QDPR; SGSH; SHMT2; SLC7A4; TYMS; ds.
XX
XX Homo sapiens.
XX
XX WO200176451-A2.
XX
```

```
PD 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP004016.
XX
XX 06-APR-2000; 2000DE-01019058.
XX
XX 07-APR-2000; 2000DE-01019173.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-010834/01.
XX
XX New nucleic acid, useful for diagnosis and therapy of metabolic disease,
XX solid tumor and cancers, comprises segment of chemically modified genomic
XX sequences of genes associated with metabolism.
XX
XX Claim 1; Page 79-84; 143pp; English.
XX
XX The invention relates to a nucleic acid (I) comprising a sequence at
XX least 18 bases of a segment of the chemically pretreated DNA of genes
XX associated with metabolism such as DUSP2 (NM_004418), EPHX2 (NM_001979),
XX QDPR (NM_000320), SGSH (NM_000199), SHMT2 (NM_005412), SLC7A2
XX (NM_003046), SLC7A4 (NM_004173) and TYMS (NM_001071) (all undefined). (I)
XX are useful for diagnosis and therapy of metabolic disease, solid tumours
XX and cancers; as primer oligonucleotides for the amplification of DNA
XX sequences; for detecting the cytosine methylation state and/or single
XX nucleotide polymorphisms (SNPs) in a chemically treated DNA of genes
XX associated with metabolism. An array of (I) is useful for ascertaining
XX genetic and/or epigenetic parameters for the diagnosis and/or therapy of
XX existing diseases or the predisposition to specific diseases by analysing
XX cytosine methylations. The method involves chemically treating genomic
XX DNA sample by a solution of bisulphite, hydrogen sulphite or disulphite
XX such that cytosine bases which are unmethylated at the 5th position are
XX converted to uracil or another base which is dissimilar to cytosine in
XX terms of hybridisation behaviour and amplifying fragments of the
XX chemically pretreated genomic DNA. The genomic DNA is from cells or
XX cellular components which contain DNA, sources of DNA comprising, for
XX e.g. cell lines, biopsies, blood, sputum, stool, urine, cerebral-spinal
XX fluid, tissue embedded in paraffin such as tissue from eye, intestine,
XX kidney, brain, heart, prostate, lung, breast or liver, histologic object
XX slides and their combinations. Genetic parameters are mutations, in
XX particular insertions, deletions, point mutations, inversions and
XX polymorphisms of genes associated with metabolism and sequences further
XX required for their regulation. Epigenetic parameters are in particular
XX cytosine methylations and further chemical modifications of DNA bases of
XX genes associated with metabolism. Further epigenetic parameters include
XX for e.g. the acetylation of histones which correlates with DNA
XX methylation. AAS63306-AAS63373 represent chemically pretreated metabolism
XX associated genes, and related primers of the invention
XX
XX Sequence 17527 BP; 4501 A; 296 C; 4094 G; 8636 T; 0 U; 0 Other;
```

```
Query Match 12.1%; Score 50.6; DB 6; Length 17527;
Best Local Similarity 52.1%; Pred. No. 0.013;
Matches 113; Conservative 0; Mismatches 104; Indels 0; Gaps 0;
QY 192 GATTTCATCACAGATAGTTCTTGTGCTAGAAATCTCTGAAATCTCGATAGTTTTCACA 251
Db 16590 GTTGTATAAGAAATATTTGTTTTTTTATTGAGTCGTATTATTAGAAAGTTTAAAGA 16649
QY 252 TGTGTAATAGATGTTCTTATTGCGGATGTTGATTAGGGTTTGTATTCTTGATTA 311
Db 16650 GGTGGTATTGTTGGTGTGTTTTTTTGTGGTTTTTTTATTATTTTTTTTAGTTT 16709
QY 312 TGGCATTCGAATAGGGAATTTCTTGGTTTGTGTGATCTTACGATACATCTCGCAA 371
Db 16710 TAATAGTGATTATTGGGAAGTTTTTTTTTGTGTTGTTTTTTTGTATTATTTTAT 16769
QY 372 TTGAATACGATCGAATCAATCTGTTAATTTGTTG 408
```





CC deficiency, viral infection, retroviral infection, Sezary syndrome,  
 CC haematological disorders, immunological disorders, Werner syndrome,  
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,  
 CC neurological disorders, neurodegenerative disorders, Waardenburg  
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
 CC infarction, hyperextension, angiogenesis, erythropoiesis, congenital heart  
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
 CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
 CC associated genomic DNA molecules of the invention. Note: The sequence  
 CC data for this patent did not form part of the printed specification but  
 CC was obtained in electronic format directly from the European Patent  
 CC Office

XX Sequence 6306 BP; 1455 A; 190 C; 1580 G; 3081 T; 0 U; 0 Other;  
 SQ

Query Match 12.0%; Score 50.2; DB 6; Length 6306;  
 Best Local Similarity 50.2%; Pred. No. 0.014;  
 Matches 124; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 161 GATTTTAGTTAGATGAATCGTATAGCTAGATTCATCACCAGATGTTCTTTGCT 220  
 |||||  
 DB 5490 GAGTTTCGGTTCGAGGGGTATAGTATGTTAGATTAATTTGTTTATAGGTAG 5549  
 |||||

QY 221 AGAATCTCTGMAATCTCGATAGTTTTCACATGCTGTAATAGATGTTCTTATTCGGCGA 280  
 |||||  
 DB 5550 ACGTTTCGGTATGATTTAGATTTTGTGTTTATGTTTATGTTTATTAAGTT 5609  
 |||||

QY 281 TTGTCGATAGGGTTTGGATTTTCTGATATGCGATTCGCAATAGGATTTCTTTGGT 340  
 |||||  
 DB 5610 TGATTTAAATTTATTTTGGTGGGTTTATATATCGTTTTTTTTTGTGTTTTTTT 5669  
 |||||

QY 341 TTGTCGTCGATCTACGATACATTCCTGCAATTTGAAATAGTATGGAATCTAAATCTTGTA 400  
 |||||  
 DB 5670 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 5729  
 |||||

QY 401 ATTGTT 407  
 |||||  
 DB 5730 TTTTTT 5736  
 |||||

RESULT 7  
 ABL33562  
 ID ABL33562 standard; DNA; 5987 BP.  
 XX  
 AC ABL33562;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 1535.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiaesthmic;  
 KW antiarteriosclerotic; anti-anaemic; cytostatic; nootropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
 ds.  
 XX  
 OS Homo sapiens.  
 XX  
 EN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP007537.  
 XX  
 PR 30-JUN-2000; 2000DE-01032529.  
 FR 01-SEP-2000; 2000DE-01043826.  
 XX  
 FA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.  
 XX Nucleic acid comprising fragment of chemically modified gene, useful for  
 XX diagnosis and treatment of diseases associated with abnormal cytosine  
 PT methylation.  
 XX Claim 1; SEQ ID NO 1535; 32pp + Sequence Listing; German.  
 XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention  
 XX

Sequence 5987 BP; 1754 A; 96 C; 1096 G; 3041 T; 0 U; 0 Other;  
 SQ

Query Match 11.7%; Score 49; DB 6; Length 5987;  
 Best Local Similarity 47.5%; Pred. No. 0.026;  
 Matches 145; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 108 TTTCTATTCTGGCCATTATCTTGTCTTCTCGCCGAATGATCTCCGTATAAGATTTTA 167  
 |||||  
 DB 2953 TTTTATTTTATGTTTATGATTTGTTATTTTAAAAATTTTGTATTTTAAAGTTTA 2912  
 |||||

QY 168 GGTAGAGATGAATCGTATAGTATTCATCACCAGATGTTCTTCTGCTAGATCT 227  
 |||||  
 DB 2913 GATTTTATTTTGAATTTTATTTTATGATTTTATGATATGGAATTTTATTTTATTT 2972  
 |||||

QY 228 CTGAAATCTCGATAGTTTTCACATGCTGTAATAGATTTCTTATTCGCGATTTGTA 287  
 |||||  
 DB 2973 GTTATATAAGTGATAGATTTTATTTATGATAGATGAGATTTAAATTTTGTAGTTTA 3032  
 |||||

QY 288 TTAGGTTTGTGATTTCTTGATATGCGATTCGCAATAGGATTTCTTTGGTTTGTGT 347  
 |||||  
 DB 3033 AAATAATTTAAATATAATTTATTTTATGTTTGTGTTTTTTTATTTTGTATTTAAAT 3092  
 |||||

QY 348 TGATCTACGATACATTCCTGCAATTCGAATGATGATCTAAATCTTGTAAATTTGTT 407  
 |||||  
 DB 3093 TTTTATTTAGATATAGATTAATTAATGTTTGTATAGGATAGATGATGATGATTTT 3152  
 |||||

QY 408 GAACA 412  
 |||||  
 DB 3153 TAAAA 3157  
 |||||

RESULT 8  
 ABQ67101  
 ID ABQ67101 standard; DNA; 5987 BP.  
 XX  
 AC ABQ67101;  
 XX  
 DT 28-AUG-2002 (first entry)  
 XX  
 DE Human angiogenesis associated polynucleotide SEQ ID NO 131.  
 XX  
 KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;  
 KW inflammation; rheumatoid arthritis; diabetic retinopathy; antilucers;  
 KW macular degeneration; inflammatory bowel disease; Crohn's disease;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiarteriosclerotic; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200246454-A2.  
 XX  
 PD 13-JUN-2002.  
 XX  
 PF 06-DEC-2001; 2001WO-EP014320.  
 XX

PR 06-DEC-2000; 2000DE-01061338.  
XX (EPIG-) EPIGENOMICS AG.  
XX Schacht O;  
XX WPI; 2002-500450/53.  
XX New nucleic acid fragments from chemically treated angiogenesis-  
PT associated genes, useful for determining methylation status, e.g. in  
PT diagnosis or treatment of cancer.  
XX Claim 1; SEQ ID NO 131; 41pp + Sequence Listing; German.  
XX The invention relates to a nucleic acid (I) comprising a segment of 18  
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)  
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also  
CC related oligomers, are used to evaluate the methylation status and/or  
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for  
CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
CC neovascular glaucoma, solid tumors, inflammation, rheumatoid arthritis,  
CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
CC Crohn's disease. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at fcp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5987 BP; 1754 A; 96 C; 1096 G; 3041 T; 0 U; 0 Other;  
Query Match 11.7%; Score 49; DB 6; Length 5987;  
Best Local Similarity 47.5%; Pred. No. 0.026;  
Matches 145; Conservative 0; Mismatches 160; Indels 0; Gaps 0;  
QY 108 TTCTATCTTCCGCAATATCTGTTCTTCCGCAATGATCTCCGATATAAGATTTA 167  
DB 2853 TTTTCTATCTTCCGCAATATCTGTTCTTCCGCAATGATCTCCGATATAAGATTTA 2912  
QY 168 GGTAGAGATGAATCGTAGCTAGATTTCATCCAGATAGTTCTTGTGCTAGATCT 227  
DB 2913 GATTTTTTTGAAATTTTTTTTTTTGATTTTAAAGTATATGGAATTTTTTATTTATTTT 2972  
QY 228 CTGAATCTCGATAGTTTTCACATGTAATAGATCTTCTTATTCGCGATGTTGA 287  
DB 2973 GTTATATAAGTAGATAGATTTTATTTATTCGGATAGATGAGTTTAAATTTTTTGAGTTTA 3032  
QY 288 TTAGGTTTGTGATTTCTTGATTCGATTCGCAATTCGGAATTTCTTGTGTTGTG 347  
DB 3033 AAATAATTTAAATATAATTTATTTTATGTTTGTGTTTATTTTATTTGATTTAAAT 3092  
QY 348 TGATCTTACGATACATCTCGCAATTCGATAGCTATGGAATCTTAATCTGTTAATTTGTT 407  
DB 3093 TTTTCTTATAGATAATAGATAATTAATGTTTGTATAGGAAGATATGATGATGATTTATTT 3152  
QY 408 GAACA 412  
DB 3153 TAAAA 3157  
RESULT 9  
ID ADB54307  
AC ADB54307 standard; DNA; 5987 BP.  
XX ADB54307;  
XX 04-DEC-2003 (first entry)  
XX Pretreated genomic DNA region 231.  
XX colon cell proliferative disorder; non methylated CpG dinucleotide;  
XX cytosinatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.  
XX Unidentified.  
XX

PN WO2003072821-A2.  
XX 04-SEP-2003.  
PD 27-FEB-2003; 2003WO-EP002035.  
PF 27-FEB-2002; 2002EP-00004551.  
PR (EPIG-) EPIGENOMICS AG.  
PA Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;  
PI Rujan T, Schmitt A;  
XX WPI; 2003-731620/59.  
XX Detecting and differentiating between colon cell proliferative disorders  
PT associated with a gene or its regulatory regions comprises contacting a  
PT target nucleic acid in a biological sample obtained from the subject with  
PT a reagent.  
XX Claim 32; SEQ ID NO 363; 74pp; English.  
XX The invention relates to a novel method for detecting and differentiating  
CC between colon cell proliferative disorders associated with at least one  
CC gene or its regulatory regions. The method comprises contacting a target  
CC nucleic acid in a biological sample obtained from the subject with at  
CC least one reagent or a series of reagents, where the reagent or series of  
CC reagents, distinguishes between methylated and non methylated CpG  
CC dinucleotides within the target nucleic acid. The molecules of the  
CC invention demonstrate cytosinatic activity whilst the method may useful  
CC for detecting and differentiating between colon cell proliferative  
CC disorders, including cancers such as colon adenoma and colon carcinoma.  
CC The PNA (peptide nucleic acid)-oligomers are useful as probes for  
CC determining cytosine methylation state or single nucleotide  
CC polymorphisms. The current sequence is that of the pretreated genomic DNA  
CC region of the invention. This sequence is not shown within the  
CC specification but is taken from Wipoweb.  
XX  
SQ Sequence 5987 BP; 1754 A; 0 C; 1096 G; 3137 T; 0 U; 0 Other;  
Query Match 11.7%; Score 49; DB 9; Length 5987;  
Best Local Similarity 47.5%; Pred. No. 0.026;  
Matches 145; Conservative 0; Mismatches 160; Indels 0; Gaps 0;  
QY 108 TTCTATCTTCCGCAATATCTGTTCTTCCGCAATGATCTCCGATATAAGATTTA 167  
DB 2853 TTTTCTATCTTCCGCAATATCTGTTCTTCCGCAATGATCTCCGATATAAGATTTA 2912  
QY 168 GGTAGAGATGAATCGTAGCTAGATTTCATCCAGATAGTTCTTGTGCTAGATCT 227  
DB 2913 GATTTTTTTGAAATTTTTTTTTTTGATTTTAAAGTATATGGAATTTTTTATTTATTTT 2972  
QY 228 CTGAATCTCGATAGTTTTCACATGTAATAGATCTTCTTATTCGCGATGTTGA 287  
DB 2973 GTTATATAAGTAGATAGATTTTATTTATTCGGATAGATGAGTTTAAATTTTTTGAGTTTA 3032  
QY 288 TTAGGTTTGTGATTTCTTGATTCGCAATTCGGAATTTCTTGTGTTGTG 347  
DB 3033 AAATAATTTAAATATAATTTATTTTATGTTTGTGTTTATTTTATTTGATTTAAAT 3092  
QY 348 TGATCTTACGATACATCTCGCAATTCGATAGCTATGGAATCTTAAATCTGTTAATTTGTT 407  
DB 3093 TTTTCTTATAGATAATAGATAATTAATGTTTGTATAGGAAGATATGATGATGATTTATTT 3152  
QY 408 GAACA 412  
DB 3153 TAAAA 3157  
RESULT 10  
ID ADB54179  
XX ADB54179 standard; DNA; 5987 BP.  
XX

AC ADB54179;  
XX  
XX  
DT 04-DEC-2003 (first entry)  
XX  
XX Pretreated genomic DNA region 103.  
XX  
XX colon cell proliferative disorder; non methylated CpG dinucleotide;  
KW cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.  
XX  
XX Unidentified.  
OS  
XX WO2003072821-A2.  
FN  
XX  
XX 04-SEP-2003.  
PD  
XX  
XX 27-FEB-2003; 2003WO-EP002035.  
XX  
XX 27-FEB-2002; 2002EP-00004551.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Adorjan P, Burger M, Maier S, Nimmrich I, Becker B, Lesche R;  
PI Rujan T, Schmitt A;  
XX  
XX WPI; 2003-731620/69.  
DR  
XX  
XX Detecting and differentiating between colon cell proliferative disorders  
PT associated with a gene or its regulatory regions comprises contacting a  
PT target nucleic acid in a biological sample obtained from the subject with  
PT a reagent.  
XX  
XX Claim 32; SEQ ID NO 235; 74pp; English.  
PS  
XX  
XX The invention relates to a novel method for detecting and differentiating  
CC between colon cell proliferative disorders associated with at least one  
CC gene or its regulatory regions. The method comprises contacting a target  
CC nucleic acid in a biological sample obtained from the subject with at  
CC least one reagent or a series of reagents, where the reagent or series of  
CC reagents, distinguishes between methylated and non methylated CpG  
CC dinucleotides within the target nucleic acid. The molecules of the  
CC invention demonstrate cytostatic activity whilst the method may useful  
CC for detecting and differentiating between colon cell proliferative  
CC disorders, including cancers such as colon adenoma and colon carcinoma.  
CC The PNA (peptide nucleic acid)-oligomers are useful as probes for  
CC determining cytosine methylation state or single nucleotide  
CC polymorphisms. The current sequence is that of the pretreated genomic DNA  
CC region of the invention. This sequence is not shown within the  
CC specification but is taken from Wipoweb.  
XX  
XX Sequence 5987 BP; 1754 A; 96 C; 1096 G; 3041 T; 0 U; 0 Other;  
SQ

Query Match 11.7%; Score 49; DB 9; Length 5987;  
Best Local Similarity 47.5%; Pred. No. 0.026;  
Matches 145; Conservative 0; Mismatches 160; Indels 0; Gaps 0;  
QY 108 TTTCATCTCTGCGCAATATCTGTTCTTCGCGCAATGATCTTCGTAATAGATTTA 167  
DB 2853 TTTTCTTTTATGTTAGATTGTTATTTTAAAAATTTTGTATTTTAAAGTTTA 2912  
QY 168 GGTTAGAGATGATCGTATAGTATTCATCACCAGATAGTTCCTTCTCTAGAAATCT 227  
DB 2913 GATTTTTCGAATTTTTTTTGTATTTTAAATATGATATGGAATTTTATTTATTT 2972  
QY 228 CTGAATCTCGATAGTTTTCACATGATGTAATAAGATGTTCTTATTCGCGCATGTTGA 287  
DB 2973 GTTATATAAGTATAGATAATTTTATTATGGATAGATGAGTTTAAATTTTTTGAAGTTA 3032  
QY 288 TTAGGGTTTTGATTTTCTTGAATATGCGATTGCAATAGGATTTTCTTTGGTTTGTGT 347  
DB 3033 AATAAATTAATAATAATTTTATGTTTGTGTTTTTTTATTTTGTATTTAAAT 3092  
QY 348 TGATCTTTACGATACATTCCTGCAATGAAATGAAATCGATCGATCTAAATCTGTTAATTTGTT 407

DB 3093 TTTTITTTAGATAGATAGATTAATTAATGTTTGTATAGGAGATAGATGATGATTATT 3152  
QY 408 GAACA 412  
DB 3153 TAAAA 3157  
RESULT 11  
ABK39964  
ID ABK39964 standard; DNA; 15479 BP.  
XX  
XX AC ABK39964;  
XX  
XX 21-MAY-2002 (first entry)  
XX Human chemically pretreated gene sequence #23 strand 1.  
DE Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;  
KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;  
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.  
XX  
XX Homo sapiens.  
OS  
XX WO200202806-A2.  
FN  
XX  
XX 10-JAN-2002.  
PD  
XX  
XX 29-JUN-2001; 2001WO-EP007470.  
PF  
XX  
XX 30-JUN-2000; 2000DE-01032529.  
PR  
XX 01-SEP-2000; 2000DE-01043826.  
PR  
XX (EPIG-) EPIGENOMICS AG.  
PA  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
PI  
XX WPI; 2002-154757/20.  
XX  
XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,  
PT useful for detecting cytosine methylation state of genes associated with  
PT pharmacogenomics and for therapy of diseases e.g. cancer.  
XX  
XX Claim 1; SEQ ID NO 45; 24pp; English.  
PS  
XX  
XX The invention relates to a nucleic acid comprising a sequence at least 18  
CC bases in length of a segment of the chemically pretreated DNA of genes  
CC associated with pharmacogenomics according to one of the sequences of the  
CC genes ALDH6 (NM 000593), CYP11A (NM 000781), CYP11B1 (NM 000497), CYP3A3  
CC (NM 000776 and NM 017460), DPYD (NM 000110), EPHX2 (NM 001979), OCLN  
CC (NM 002538), TXNRD1 (NM 003330), UGT8 (NM 003360), MRP (NM 004996,  
CC NM 019900, NM 019901, NM 019902, NM 019862, NM 019898, NM 019899) and  
CC their complementary sequences, or a sequence (S1) chosen from 87  
CC sequences and their complements. The chemical pretreatment is bisulphite  
CC treatment to convert cytosines (but not methyl-cytosines) into uracil.  
CC Also included are an oligomer (II) in particular an oligonucleotide or a  
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one  
CC base sequence having a length of 9 nucleotides which hybridises to or is  
CC identical to a chemically pretreated DNA of genes associated with  
CC pharmacogenomics and their complements, arranged in an array for  
CC analysing diseases associated with the methylation state (CpG) and/or  
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The  
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids  
CC and their complements is useful for diagnosis and therapy of solid  
CC tumours and cancer. The present sequence represents one the 87 DNA  
CC sequences or its complement. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 15479 BP; 4303 A; 122 C; 2850 G; 8204 T; 0 U; 0 Other;  
SQ

Query Match 11.7%; Score 48.8; DB 6; Length 15479;  
Best Local Similarity 51.9%; Pred. No. 0.034;

[illegible][illegible]

RESULT 13  
ABK39976  
ID ABK39976 standard: DNA: 17848 BP.

AA ABK39976;

DT 21-MAY-2002 (first entry)

DE Human chemically pretreated gene sequence #29 strand 2.  
yy

KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; TXNRD1; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; CYP2C8; CYP2C9; CYP2C19; CYP2D6; CYP2E1; CYP3A4; CYP3A5; CYP3A7; CYP3A9; CYP3A10; CYP3A11; CYP3A12; CYP3A13; CYP3A14; CYP3A15; CYP3A16; CYP3A17; CYP3A18; CYP3A19; CYP3A20; CYP3A21; CYP3A22; CYP3A23; CYP3A24; CYP3A25; CYP3A26; CYP3A27; CYP3A28; CYP3A29; CYP3A30; CYP3A31; CYP3A32; CYP3A33; CYP3A34; CYP3A35; CYP3A36; CYP3A37; CYP3A38; CYP3A39; CYP3A40; CYP3A41; CYP3A42; CYP3A43; CYP3A44; CYP3A45; CYP3A46; CYP3A47; CYP3A48; CYP3A49; CYP3A50; CYP3A51; CYP3A52; CYP3A53; CYP3A54; CYP3A55; CYP3A56; CYP3A57; CYP3A58; CYP3A59; CYP3A60; CYP3A61; CYP3A62; CYP3A63; CYP3A64; CYP3A65; CYP3A66; CYP3A67; CYP3A68; CYP3A69; CYP3A70; CYP3A71; CYP3A72; CYP3A73; CYP3A74; CYP3A75; CYP3A76; CYP3A77; CYP3A78; CYP3A79; CYP3A80; CYP3A81; CYP3A82; CYP3A83; CYP3A84; CYP3A85; CYP3A86; CYP3A87; CYP3A88; CYP3A89; CYP3A90; CYP3A91; CYP3A92; CYP3A93; CYP3A94; CYP3A95; CYP3A96; CYP3A97; CYP3A98; CYP3A99; CYP3A100; CYP3A101; CYP3A102; CYP3A103; CYP3A104; CYP3A105; CYP3A106; CYP3A107; CYP3A108; CYP3A109; CYP3A110; CYP3A111; CYP3A112; CYP3A113; CYP3A114; CYP3A115; CYP3A116; CYP3A117; CYP3A118; CYP3A119; CYP3A120; CYP3A121; CYP3A122; CYP3A123; CYP3A124; CYP3A125; CYP3A126; CYP3A127; CYP3A128; CYP3A129; CYP3A130; CYP3A131; CYP3A132; CYP3A133; CYP3A134; CYP3A135; CYP3A136; CYP3A137; CYP3A138; CYP3A139; CYP3A140; CYP3A141; CYP3A142; CYP3A143; CYP3A144; CYP3A145; CYP3A146; CYP3A147; CYP3A148; CYP3A149; CYP3A150; CYP3A151; CYP3A152; CYP3A153; CYP3A154; CYP3A155; CYP3A156; CYP3A157; CYP3A158; CYP3A159; CYP3A160; CYP3A161; CYP3A162; CYP3A163; CYP3A164; CYP3A165; CYP3A166; CYP3A167; CYP3A168; CYP3A169; CYP3A170; CYP3A171; CYP3A172; CYP3A173; CYP3A174; CYP3A175; CYP3A176; CYP3A177; CYP3A178; CYP3A179; CYP3A180; CYP3A181; CYP3A182; CYP3A183; CYP3A184; CYP3A185; CYP3A186; CYP3A187; CYP3A188; CYP3A189; CYP3A190; CYP3A191; CYP3A192; CYP3A193; CYP3A194; CYP3A195; CYP3A196; CYP3A197; CYP3A198; CYP3A199; CYP3A200; CYP3A201; CYP3A202; CYP3A203; CYP3A204; CYP3A205; CYP3A206; CYP3A207; CYP3A208; CYP3A209; CYP3A210; CYP3A211; CYP3A212; CYP3A213; CYP3A214; CYP3A215; CYP3A216; CYP3A217; CYP3A218; CYP3A219; CYP3A220; CYP3A221; CYP3A222; CYP3A223; CYP3A224; CYP3A225; CYP3A226; CYP3A227; CYP3A228; CYP3A229; CYP3A230; CYP3A231; CYP3A232; CYP3A233; CYP3A234; CYP3A235; CYP3A236; CYP3A237; CYP3A238; CYP3A239; CYP3A240; CYP3A241; CYP3A242; CYP3A243; CYP3A244; CYP3A245; CYP3A246; CYP3A247; CYP3A248; CYP3A249; CYP3A250; CYP3A251; CYP3A252; CYP3A253; CYP3A254; CYP3A255; CYP3A256; CYP3A257; CYP3A258; CYP3A259; CYP3A260; CYP3A261; CYP3A262; CYP3A263; CYP3A264; CYP3A265; CYP3A266; CYP3A267; CYP3A268; CYP3A269; CYP3A270; CYP3A271; CYP3A272; CYP3A273; CYP3A274; CYP3A275; CYP3A276; CYP3A277; CYP3A278; CYP3A279; CYP3A280; CYP3A281; CYP3A282; CYP3A283; CYP3A284; CYP3A285; CYP3A286; CYP3A287; CYP3A288; CYP3A289; CYP3A290; CYP3A291; CYP3A292; CYP3A293; CYP3A294; CYP3A295; CYP3A296; CYP3A297; CYP3A298; CYP3A299; CYP3A300; CYP3A301; CYP3A302; CYP3A303; CYP3A304; CYP3A305; CYP3A306; CYP3A307; CYP3A308; CYP3A309; CYP3A310; CYP3A311; CYP3A312; CYP3A313; CYP3A314; CYP3A315; CYP3A316; CYP3A317; CYP3A318; CYP3A319; CYP3A320; CYP3A321; CYP3A322; CYP3A323; CYP3A324; CYP3A325; CYP3A326; CYP3A327; CYP3A328; CYP3A329; CYP3A330; CYP3A331; CYP3A332; CYP3A333; CYP3A334; CYP3A335; CYP3A336; CYP3A337; CYP3A338; CYP3A339; CYP3A340; CYP3A341; CYP3A342; CYP3A343; CYP3A344; CYP3A345; CYP3A346; CYP3A347; CYP3A348; CYP3A349; CYP3A350; CYP3A351; CYP3A352; CYP3A353; CYP3A354; CYP3A355; CYP3A356; CYP3A357; CYP3A358; CYP3A359; CYP3A360; CYP3A361; CYP3A362; CYP3A363; CYP3A364; CYP3A365; CYP3A366; CYP3A367; CYP3A368; CYP3A369; CYP3A370; CYP3A371; CYP3A372; CYP3A373; CYP3A374; CYP3A375; CYP3A376; CYP3A377; CYP3A378; CYP3A379; CYP3A380; CYP3A381; CYP3A382; CYP3A383; CYP3A384; CYP3A385; CYP3A386; CYP3A387; CYP3A388; CYP3A389; CYP3A390; CYP3A391; CYP3A392; CYP3A393; CYP3A394; CYP3A395; CYP3A396; CYP3A397; CYP3A398; CYP3A399; CYP3A400; CYP3A401; CYP3A402; CYP3A403; CYP3A404; CYP3A405; CYP3A406; CYP3A407; CYP3A408; CYP3A409; CYP3A410; CYP3A411; CYP3A412; CYP3A413; CYP3A414; CYP3A415; CYP3A416; CYP3A417; CYP3A418; CYP3A419; CYP3A420; CYP3A421; CYP3A422; CYP3A423; CYP3A424; CYP3A425; CYP3A426; CYP3A427; CYP3A428; CYP3A429; CYP3A430; CYP3A431; CYP3A432; CYP3A433; CYP3A434; CYP3A435; CYP3A436; CYP3A437; CYP3A438; CYP3A439; CYP3A440; CYP3A441; CYP3A442; CYP3A443; CYP3A444; CYP3A445; CYP3A446; CYP3A447; CYP3A448; CYP3A449; CYP3A450; CYP3A451; CYP3A452; CYP3A453; CYP3A454; CYP3A455; CYP3A456; CYP3A457; CYP3A458; CYP3A459; CYP3A460; CYP3A461; CYP3A462; CYP3A463; CYP3A464; CYP3A465; CYP3A466; CYP3A467; CYP3A468; CYP3A469; CYP3A470; CYP3A471; CYP3A472; CYP3A473; CYP3A474; CYP3A475; CYP3A476; CYP3A477; CYP3A478; CYP3A479; CYP3A480; CYP3A481; CYP3A482; CYP3A483; CYP3A484; CYP3A485; CYP3A486; CYP3A487; CYP3A488; CYP3A489; CYP3A490; CYP3A491; CYP3A492; CYP3A493; CYP3A494; CYP3A495; CYP3A496; CYP3A497; CYP3A498; CYP3A499; CYP3A500; CYP3A501; CYP3A502; CYP3A503; CYP3A504; CYP3A505; CYP3A506; CYP3A507; CYP3A508; CYP3A509; CYP3A510; CYP3A511; CYP3A512; CYP3A513; CYP3A514; CYP3A515; CYP3A516; CYP3A517; CYP3A518;

KW KWK; pharmacogenomics; SAE; single nucleotide polymorphism  
XX

[illegible][illegible]

XX  
PF 29-JUN-2001: 2001WO-EP007470.

XX  
PR 30-JUN-2000; 2000DE-01032529.

XX  
v7

XX  
XX

XX  
DP WPT: 2002-154757/20.

XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,  
PT

pharmacogenomics and for therapy of diseases e.g. cancer.

PS Claim 1; SEQ ID NO 58; 24pp; English.

CC The invention relates to a nucleic acid comprising a sequence at least 18  
 CC bases in length of a segment of the chemically pretreated DNA of genes  
 CC associated with pharmacogenomics according to one of the sequences of the  
 CC genes ALDH6 (NM 000693), CYP11A (NM 000781), CYP11B (NM 000497), CYP3A3  
 CC (NM 000776 and NM 017460), DPYD (NM 000110), EPHX2 (NM 001979), OCLN  
 CC (NM 002538), TXNRD1 (NM 003330), UGT8 (NM 003360), MRP (NM 004996,  
 CC NM 019900, NM 019901, NM 019902, NM 019862, NM 019898, NM 019899) and  
 CC their complementary sequences, or a sequence (S<sub>i</sub>) chosen from 87  
 CC sequences and their complements. The chemical pretreatment is bisulphite  
 CC treatment to convert cytosines (but not methyl-cytosines) into uracils.  
 CC Also included are an oligomer (II) in particular an oligonucleotide or a  
 CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one  
 CC base sequence having a length of 9 nucleotides which hybridises to or is  
 CC identical to a chemically pretreated DNA of genes associated with  
 CC pharmacogenomics and their complements, arranged in an array for  
 CC analysing diseases associated with the methylation state (CpG) and/or  
 CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The  
 CC oligomers may also be used as PCR primers. The set of 87 nucleic acids  
 CC and their complements is useful for diagnosis and therapy of solid  
 CC tumours and cancer. The present sequence represents one the 87 DNA  
 CC sequences or its complement. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 0 U; 6 Other;

Query Match 11.6%; Score 48.4; DB 6; Length 17848;  
 Best Local Similarity 46.9%; Pred. No. 0.044;  
 Matches 151; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 91 AAATTAGGCGAGATTAGTTCTTATCTTGGCCATATCTGTTCTTCCGCCAATGATCT 150  
 DB 1226 ABAATATTTAAAGTTAATATATTTTAAATGAATTAATTTTATTATTAATTCAG 1285

QY 151 TCGTATAAGATTTAGGTTAGATGAATCGTATAGTAGATTTTACATGCTGAATGATGTTCT 270  
 DB 1286 TTTTCTTTGTTAAATTTAGTGGTTATTAATTTTCTGTTTGTATTTTATTTTATA 1405

QY 271 TATTCGGCGATTGTTAGTTAGGTTTGTATTTCTTGTATTCGATTCGCAATAGGAT 330  
 DB 1406 TATTTGTTTCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1465

QY 331 TTTCTTTGGTTTGTGTTGATCTTACGATACATTCCTGCAATGATAGTATGATCTA 390  
 DB 1466 TGTCTTTTTCGATGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 1525

QY 391 AATCTGTTTAAATTTGTTGACA 412  
 DB 1526 ATATGAATAATATTTGTTTATA 1547

RESULT 14  
 ABK28164  
 ID ABK28164 standard; DNA; 17848 BP.

XX ABK28164;

XX 23-APR-2002 (first entry)

XX DNA transcription associated complementary genomic DNA #19.

XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
 KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
 KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
 KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
 KW immunological disorder; Werner syndrome; developmental disorder;  
 KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;  
 KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;

KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;  
 KW angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;  
 KW polyglutamine disorder; solid tumour.

XX Unidentified.

XX WO2001192565-A2.

XX 06-DEC-2001.

XX 06-APR-2001; 2001WO-EP003973.

XX 06-APR-2000; 2000DE-01019058.

XX 07-APR-2000; 2000DE-01019173.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPICENOVICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-090046/12.

XX New nucleic acids or oligomers, useful for diagnosing or treating  
 XX diseases associated with DNA transcription, e.g. immunological disorders,  
 XX Werner syndrome, psoriasis, myocardial infarction, solid tumors or  
 XX cancer.

XX Claim 1; SEQ ID NO 38; 32pp; English.

XX The invention relates to a nucleic acid, which comprises a segment of the  
 XX chemically pretreated DNA of genes associated with DNA transcription from  
 XX one of 346 sequences, and an oligomer, in particular an oligonucleotide  
 XX or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
 XX to the chemically pretreated DNA of genes associated with DNA  
 XX transcription. The set of oligomer probes are useful for detecting the  
 XX cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
 XX in a chemically pretreated genomic DNA. The nucleic acids are useful for  
 XX diagnosing or treating diseases associated with DNA transcription  
 XX (particularly with the methylation status), e.g. adenosine deaminase  
 XX deficiency, viral infection, retroviral infection, Sezary syndrome,  
 XX haematological disorders, immunological disorders, Werner syndrome,  
 XX tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,  
 XX neurological disorders, neurodegenerative disorders, Waardenburg  
 XX syndrome, Niemann-Pick disease, myelodysplastic syndrome, congenital heart  
 XX infarction, hypertension, angiodysplasia, erythropoiesis, solid tumours  
 XX disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
 XX or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
 XX associated genomic DNA molecules of the invention. Note: The sequence  
 XX data for this patent did not form part of the printed specification but  
 XX was obtained in electronic format directly from the European Patent  
 XX Office

XX Sequence 17848 BP; 5055 A; 211 C; 3533 G; 9043 T; 0 U; 6 Other;

Query Match 11.6%; Score 48.4; DB 6; Length 17848;  
 Best Local Similarity 46.9%; Pred. No. 0.044;  
 Matches 151; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 91 AAATTAGGCGAGATTAGTTCTTATCTTGGCCATATCTGTTCTTCCGCCAATGATCT 150

DB 1226 ABAATATTTAAAGTTAATATATTTTAAATGAATTAATTTTATTATTAATTCAG 1285

QY 151 TCGTATAAGATTTTAGGTTAGATGAATCGTATAGTAGATTTTACATCACCAGATAGT 210

DB 1286 TTTTCTTTGTTAAATTTAGTGGTTATTAATTTTCTGTTTGTATTTTATTTATTAAGATT 1345

QY 211 TTTCTTTGTTAGATCTCTGAATTTCTCGATAGTTTTCACATGCTGAATAGATTGTTCT 270

DB 1346 TTTTCTTTGTTAAATTTTAGTGGTTATTAATTTTCTGTTTGTATTTTATTTATTTTATA 1405

QY 271 TATTCGGCGATTGTTAGTGGGTTTGTATTTCTGTTATTCGCAATGCAATAGGAT 330

```

SQ   Sequence 2501 BP; 532 A; 124 C; 707 G; 1138 T; 0 U; 0 Other;
      Query Match      11.5%; Score 48; DB 7; Length 2501;
      Best Local Similarity 47.1%; Pred. No. 0.04;
      Matches 180; Conservative 0; Mismatches 200; Indels 2; Gaps 1;

QY   25  TTGATTTTCTCGAAATATTTTTTCGGTGATCGTGAAACTACTCGAATCGCTCGATAGGT 84

Db    1661  TTTATATTTTTCGAGATTAGTTTTCGGTTAAATAGTTTGAATTTAAATCTGCAAGGGA 1720

```

D5 1721 AAAAGGAAGTTGTTGTTATATTCGTTTTTTTGTGTTTTAGTTTTTTAAATTGTGAAA 1780

|      |     |   |      |
|------|-----|---|------|
| 1762 | 139 | AGAGATTTCTGATGTTTATTTATTTCTGGAGATGAGGAGGATTTTATATAGGTTTTAGATTTTAT | 184  |
| 205  | QY  | GATGATTTCTTTTGTCTAGAAATCTCTGAANCTTCGATAGTATTTTCCACATGTGTAAATAGAT  | 264  |
| 1841 | Db  | AAATAGTTGTGTAGTTATTATTATGATGATGATTAACGATATGTTATTTTGTGGGAAAA       | 1900 |
| 265  | QY  | TGTTCTTATTCGGCGGATGTTGATTAGGTTTTTGATTTTCTTGATTAATCGCAATTGCAATT    | 324  |
| 1901 | Db  | TAATATTTTTTGTGTAGTAGTAGTATTCGTTTTTTTTTTTTTTT-TCGTATTATTATAAATT    | 1958 |

```

1959 TTGGTTTATTATTTATTTGGTTTAGTTTTTTTAGGTGATTTCCTTGAGGGGTGA 2018
QY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
385 GATCTAAATCTGTGTAATTGTT 406
QY      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2019 GATATTATTGTTTTTGTTTTT 2040
Search completed: June 20, 2004, 01:58:56
```

Search completed: June 20, 2004, 01:58:56  
Job time : 312.833 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 01:39:08 ; Search time 2071.67 Seconds  
(without alignments)  
6025.290 Million cell updates/sec

Title: US-09-000-062-6  
Perfect score: 418  
Sequence: 1 tggagtagcattcttcgac.....taattttgtaacagatccc 418

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estinu:\*  
4: em\_estinu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hccl:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_hccl:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_pig:\*  
27: em\_gss\_vrl:\*  
28: gb\_gssl:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| C 1        | 379.8 | 90.9        | 436    | 28    | BH633573    |
| C 2        | 353.2 | 84.5        | 366    | 28    | BH846820    |
| C 3        | 194.4 | 46.5        | 258    | 29    | AL950572    |
| C 4        | 122.4 | 46.0        | 262    | 29    | AL950573    |

|   |    |      |      |      |    |          |           |
|---|----|------|------|------|----|----------|-----------|
| C | 5  | 66   | 15.8 | 223  | 29 | AL950952 | Arabidops |
| C | 6  | 56.6 | 13.5 | 1180 | 13 | BX436369 | Arabidops |
| C | 7  | 54.6 | 13.1 | 1180 | 9  | AL565108 | AL565108  |
| C | 8  | 54.4 | 13.0 | 750  | 29 | CNS0111D | AL100303  |
| C | 9  | 53.4 | 12.8 | 1201 | 9  | AL524807 | AL524807  |
| C | 10 | 52.6 | 12.6 | 960  | 29 | CNS005F3 | AL059925  |
| C | 11 | 52.4 | 12.5 | 832  | 29 | CNS0081L | Drosophil |
| C | 12 | 51.8 | 12.4 | 712  | 13 | BX416727 | BX416727  |
| C | 13 | 51.6 | 12.3 | 1101 | 29 | CNS0026Z | AL097301  |
| C | 14 | 51.6 | 12.0 | 994  | 29 | CNS018BG | Drosophil |
| C | 15 | 50.2 | 12.0 | 975  | 29 | CNS014RT | AL104531  |
| C | 16 | 50   | 12.0 | 1020 | 29 | CNS038G9 | AL232578  |
| C | 17 | 49.8 | 11.9 | 878  | 29 | CNS0187R | Tetraodon |
| C | 18 | 49.8 | 11.9 | 1124 | 13 | BX436282 | AL108993  |
| C | 19 | 49.4 | 11.8 | 953  | 13 | BX331505 | Drosophil |
| C | 20 | 49.2 | 11.8 | 524  | 29 | CNS01090 | BX331505  |
| C | 21 | 49.2 | 11.8 | 997  | 13 | BX419578 | AL167541  |
| C | 22 | 48.8 | 11.7 | 783  | 29 | CNS00A1S | Tetraodon |
| C | 23 | 48.8 | 11.7 | 1015 | 13 | BX399782 | BX419578  |
| C | 24 | 48.6 | 11.6 | 843  | 29 | AG177964 | AL055833  |
| C | 25 | 48.4 | 11.6 | 695  | 28 | BZ462901 | Drosophil |
| C | 26 | 48.4 | 11.6 | 837  | 29 | CNS07ABZ | AG177964  |
| C | 27 | 48.4 | 11.6 | 1201 | 13 | BX403954 | BX399782  |
| C | 28 | 48.2 | 11.5 | 888  | 29 | CNS00319 | AL436389  |
| C | 29 | 48.2 | 11.5 | 964  | 29 | CNS006N9 | T7 end of |
| C | 30 | 48.2 | 11.5 | 1101 | 29 | CNS00LJT | AL436389  |
| C | 31 | 48.2 | 11.5 | 1201 | 13 | BX446430 | AL436389  |
| C | 32 | 48   | 11.5 | 759  | 29 | CNS060XV | AL411257  |
| C | 33 | 48   | 11.5 | 1043 | 29 | CNS0145P | T7 end of |
| C | 34 | 48   | 11.5 | 1157 | 29 | CNS020GV | AL103735  |
| C | 35 | 48   | 11.5 | 1201 | 9  | AL554111 | Drosophil |
| C | 36 | 47.8 | 11.4 | 1101 | 29 | CNS008XS | AL214456  |
| C | 37 | 47.8 | 11.4 | 1101 | 29 | CNS012NV | Tetraodon |
| C | 38 | 47.6 | 11.4 | 1068 | 29 | CNS00J4X | AL052569  |
| C | 39 | 47.6 | 11.4 | 1101 | 29 | CNS00D7M | Drosophil |
| C | 40 | 47.6 | 11.4 | 1101 | 29 | CNS00EVL | AL075933  |
| C | 41 | 47.4 | 11.3 | 996  | 29 | CNS00DIX | AL060328  |
| C | 42 | 47.4 | 11.3 | 1165 | 13 | BX338369 | AL069706  |
| C | 43 | 47.2 | 11.3 | 1047 | 28 | CC266341 | Drosophil |
| C | 44 | 47.2 | 11.3 | 1061 | 13 | BX437039 | BX338369  |
| C | 45 | 47.2 | 11.3 | 1101 | 29 | CNS00B37 | CC266341  |

ALIGNMENTS

RESULT 1  
BH633573/c 436 bp DNA linear GSS 01-FEB-2002  
LOCUS  
DEFINITION  
SALK\_042781 Arabidopsis thaliana T-DNA insertion lines Arabidopsis thaliana genomic clone SALK\_042781, genomic survey sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BH633573.1 GI:18468936  
GSS.  
Arabidopsis thaliana (Chale cress)

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE  
AUTHORS

Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome  
Unpublished (2001)  
Contact: Joseph R. Ecker

JOURNAL  
COMMENT

Salk Institute Genomic Analysis Laboratory (SIGAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 556 6379  
Email: ecker@salk.edu

This is single pass sequence recovered from the left border of  
TDNA. This sequence lies within an annotated exon of At4g40040.  
Class: TDNA tagged

## FEATURES

Location/Qualifiers  
1. 436  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="SALK\_042781"  
/clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

## ORIGIN

Query Match 90.9%; Score 379.8; DB 28; Length 436;  
Best Local Similarity 99.5%; Pred. No. 1.6e-67;  
Matches 381; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 33 CCTGGAAATATTTTCGGTATCGTGAATCTGGAATCGCTCGATAGGTGTCAGAA 92  
DB 435 CCTGGAAATATTTTCGGTATCGTGAATCTGGAATCGCTCGATAGGTGTCAGAA 376  
QY 93 ATTAGGCGAGATTAGTTCTATTCTTGGCCATTATCTGTTCTTCCCGAATGATCTTC 152  
DB 375 ATTAGGCGAGATTAGTTCTATTCTTGGCCATTATCTGTTCTTCCCGAATGATCTTC 316  
QY 153 CGTATAAGATTTAGGTTAGAGATGAATCGTATAGCTAGATTTTCATCACCAGATGTTT 212  
DB 315 CGATAAAGATTTAGGTTAGAGATGAATCGTATAGCTAGATTTTCATCACCAGATGTTT 256  
QY 213 CTTTGTCTAGATCTCTGAATTCCTGATAGTTTTCATGATGTAATAGATTCTTCTTA 272  
DB 255 CTTTGTCTAGATCTCTGAATTCCTGATAGTTTTCATGATGTAATAGATTCTTCTTA 196  
QY 273 TTCGCGGATTTGATAGGTTTGTATTTCTTGTATTCGATTCGATTCGAAATAGGATTT 332  
DB 195 TTCGCGGATTTGATAGGTTTGTATTTCTTGTATTCGATTCGATTCGAAATAGGATTT 136  
QY 333 TCTTTGGTTTGTGTGTATCTTACGATACATCTCCGCAATTCGATGATGATCTAA 392  
DB 135 TCTTTGGTTTGTGTGTATCTTACGATACATCTCCGCAATTCGATGATGATCTAA 76  
QY 393 TCTTTGTTAATTTGTGAACAGAT 415  
DB 75 TCTTTGTTAATTTGTGAACAGAT 53

RESULT 2  
BH46820/c 366 bp DNA linear GSS 13-JUN-2002  
LOCUS SALK\_010583.54.50.x Arabidopsis thaliana TDNA insertion lines  
DEFINITION Arabidopsis thaliana genomic clone SALK\_010583.54.50.x, genomic  
survey sequence.

ACCESSION BH46820

VERSION BH46820.1

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 366)

REFERENCE

Alonso J.M., Leisse T.J., Barajas P., Chen H., Cheuk R.,  
Gadrianab C., Jeske A., Karnes M., Kim C.J., Parker H., Prednis L.,  
Shinn P., Zimmerman J. and Ecker J.R.

AUTHORS

A Sequence-Indexed Library of Insertion Mutations in the

Arabidopsis Genome

Unpublished (2001)

## COMMENT

Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGnAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: ecker@salk.edu  
This is single pass sequence recovered from the left border of  
TDNA. This sequence lies within 300 bases of the 5' end of  
At4g40040.  
Class: TDNA tagged.

## Location/Qualifiers

## 1. 366

## /organism="Arabidopsis thaliana"

## /mol\_type="genomic DNA"

## /strain="Columbia 0"

## /db\_xref="taxon:3702"

## /clone="SALK\_010583.54.50.x"

## /clone\_lib="Arabidopsis thaliana TDNA insertion lines"

## /note="PCR was performed on Arabidopsis thaliana lines

## each of which contains one or more TDNA insertion

## elements. The resultant fragment for each line was

## directly sequenced to determine the genomic sequence at

## the site of insertion. Details of the protocols used can

be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"

## ORIGIN

Query Match 84.5%; Score 353.2; DB 28; Length 366;  
Best Local Similarity 97.8%; Pred. No. 4.1e-62;  
Matches 358; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
QY 30 TTCTCTGGAAATATTTTTCGGTATCGTGAATCTGGAATCGCTCGATAGGTGTCAC 89  
DB 366 TTCTCTGGAAATATTTTTCGGTATCGTGAATCTGGAATCGCTCGATAGGTGTCAC 307  
QY 90 GAAATTTAGCGGAGATTAGTTTCTATTCTTGGCAATATCTTCTTTCGCCGATGATC 149  
DB 306 GAAATTTAGCGGAGATTAGTTTCTATTCTTGGCAATATCTTCTTTCGCCGATGATC 247  
QY 150 TTCCGTATTAAGATTTTAGGTTAGAGATGAATTCGTATAGCTAGATTTTCATCACCAGATAG 209  
DB 246 TTCCGTATTAAGATTTTAGGTTAGAGATGAATTCGTATAGCTAGATTTTCATCACCAGATAG 187  
QY 210 TTCTTTTGTCTAGATCTCTGAAATCTCGATAGTTTTCACATGTTTCAATAGATTGTC 269  
DB 186 TTCTTTTGTCTAGATCTCTGAAATCTCGATAGTTTTCACATGTTTCAATAGATTGTC 127  
QY 270 TTATTCGGCGATTGTTGATTAGGTTTTCGATTTTCTTGTATTCGATTCGCAATTAGGGA 329  
DB 126 TTATTCGGCGATTGTTGATTAGGTTTTCGATTTTCTTGTATTCGCAATTAGGGA 67  
QY 330 TTCTCTTTCGTTTGTGTTGATCTTACGATACATCTTCGCAATTCGATGATGATCT 389  
DB 66 TTCTCTTTCGTTTGTGTTGATCTTACGATACATCTTCGCAATTCGATGATGATGAT 7  
QY 390 AAATCT 395  
DB 6 AAATCT 1

## RESULT 3

AL950572/c

LOCUS

DEFINITION

Arabidopsis thaliana T-DNA flanking sequence GK-329E10-016035,

genomic survey sequence.

ACCESSION

AL950572

VERSION

AL950572.1

KEYWORDS

GSS.

SOURCE

Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

```

REFERENCE
AUTHORS      Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Siedler,H.
              and Weisshaar,B.
TITLE        A pipeline for automated high-throughput generation of FSTs
              (flanking sequence tags) from Arabidopsis thaliana T-DNA
              transformed lines
JOURNAL
REFERENCE     Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
AUTHORS      Li,Y., Strizhov,N., Rosso,M. and Weisshaar,B.
TITLE        Submitted (21-OCT-2002) Weisshaar B., Max-Planck-Institut fuer
              Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
              This sequence is recovered from the left border of the T-DNA. It
              indicates an insertion close to or within gene At4g40040. The
              sequences are generated at the MPI for Plant Breeding Research in
              the context of the GABI-Kat project. GABI-Kat is part of the German
              Plant Genomics program designated 'GABI'. Information on line
              availability can be found at:
              http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES
source       Location/Qualifiers
              1..258
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /strain="Columbia 0"
                /db_xref="taxon:3702"
                /clone="GK-329E10-016035"
                /notes="PCR was performed on DNA from Arabidopsis thaliana
                plants (T1) which were transformed with the T-DNA from
                vector pAC161. The lines contain one or more T-DNA
                insertions. The DNA fragment(s) resulting from the PCR
                were directly sequenced to determine the genomic sequence
                flanking the insertion. Sequences displaying significant
                similarity to the A. thaliana nuclear genome sequence were
                processed for submission. T-DNA derived sequences were
                removed"
ORIGIN
Query Match 46.5%; Score 194.4; DB 29; Length 258;
Best Local Similarity 99.5%; Pred. No. 8.4e-30;
Matches 195; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 220 TAGAATCTCTGAAATCTCGATAGTTTTCACATGTGTAATAGATTGTTCTTATTCGGCG 279
Db 236 TAGAATCTCTGAAATCTCGATAGTTTTCACATGTGTAATAGATTGTTCTTATTCGGCG 177
QY 280 ATTGTTGATTAGGTTTTCGATAGTTTTCGATAGTTGCGATTGCGAATTAGGATTTCCTTGG 339
Db 176 ATTGTTGATTAGGTTTTCGATAGTTTTCGATAGTTGCGATTGCGAATTAGGATTTCCTTGG 117
QY 340 TTTTGTGTTGATCTTACGATACATCTCTGCAATGTAATGATCGATCTTAATCTTGT 399
Db 116 TTTTGTGTTGATCTTACGATACATCTCTGCAATGTAATGATCGATCTTAATCTTGT 57
QY 400 AATTGTTGTAACAGAT 415
Db 56 AATTGTTGTAACAGAT 41
RESULT 4
AL950573/c
LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-329E10-016044,
DEFINITION Arabidopsis thaliana T-DNA flanking sequence.
ACCESSION AL950573
VERSION AL950573.1 GI:24407195
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)

```

```

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Siedler,H.
  and Weisshaar,B.
  A pipeline for automated high-throughput generation of FSTs
  (flanking sequence tags) from Arabidopsis thaliana T-DNA
  transformed lines
2 Unpublished
3 Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.
  A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)
  for flanking sequence tag based reverse genetics
4 Unpublished
5 Strizhov,N., Li,Y., Rosso,M. and Weisshaar,B.
  Direct Submission
  Submitted (21-OCT-2002) Weisshaar B., Max-Planck-Institut fuer
  Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany
  This sequence is recovered from the left border of the T-DNA. It
  indicates an insertion close to or within gene At4g40040. The
  sequences are generated at the MPI for Plant Breeding Research in
  the context of the GABI-Kat project. GABI-Kat is part of the German
  Plant Genomics program designated 'GABI'. Information on line
  availability can be found at:
  http://www.mpiz-koeln.mpg.de/GABI-Kat/.
FEATURES
source       Location/Qualifiers
              1..262
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /strain="Columbia 0"
                /db_xref="taxon:3702"
                /clone="GK-329E10-016044"
                /notes="PCR was performed on DNA from Arabidopsis thaliana
                plants (T1) which were transformed with the T-DNA from
                vector pAC161. The lines contain one or more T-DNA
                insertions. The DNA fragment(s) resulting from the PCR
                were directly sequenced to determine the genomic sequence
                flanking the insertion. Sequences displaying significant
                similarity to the A. thaliana nuclear genome sequence were
                processed for submission. T-DNA derived sequences were
                removed"
ORIGIN
Query Match 46.0%; Score 192.4; DB 29; Length 262;
Best Local Similarity 99.5%; Pred. No. 2.1e-29;
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 220 TAGAATCTCTGAAATCTCGATAGTTTTCACATGTGTAATAGATTGTTCTTATTCGGCG 279
Db 262 TAGAATCTCTGAAATCTCGATAGTTTTCACATGTGTAATAGATTGTTCTTATTCGGCG 203
QY 280 ATTGTTGATTAGGTTTTCGATAGTTTTCGATAGTTGCGATTGCGAATTAGGATTTCCTTGG 339
Db 202 ATTGTTGATTAGGTTTTCGATAGTTTTCGATAGTTGCGATTGCGAATTAGGATTTCCTTGG 143
QY 340 TTTTGTGTTGATCTTACGATACATCTCTGCAATGTAATGATCGATCTTAATCTTGT 399
Db 142 TTTTGTGTTGATCTTACGATACATCTCTGCAATGTAATGATCGATCTTAATCTTGT 83
QY 400 AATTGTTGTAACAG 413
Db 82 AATTGTTGTAACAG 69
RESULT 5
AL950952/c
LOCUS Arabidopsis thaliana T-DNA flanking sequence GK-332E10-016059,
DEFINITION Arabidopsis thaliana T-DNA flanking sequence.

```





Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On Feb 13, 2001 this sequence version replaced gi:12788300.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by life technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 3928.f. For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DC008AF01NP1&cluster=3928.f. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DC008AF01NP1.  
Location/Qualifiers  
1..1201

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

FEATURES  
source

/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DC008YK01"  
/tissue\_type="NEUROBLASTOMA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 12.8%; Score 53.4; DB 9; Length 1201;  
Best Local Similarity 42.9%; Pred. No. 0.3;  
Matches 129; Conservative 23; Mismatches 149; Indels 0; Gaps 0;  
QY 104 TTAGTTCTTATCTGGCATTATCTCTTCTTCGCGGAATGATCTTCGGTATAAAGAT 163  
DB 203 TTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 262  
QY 164 TTAGTTAGAGATGAATCGTATAGTATGATTCATACACGAGATGTTCTTTGTTT 223  
DB 263 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 322  
QY 224 ATCTCTGAATTCCTGATAGTTTTCACATGTTGAATAGTTTCTTTATTCGCGGATTG 283  
DB 323 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 382  
QY 284 TTGATTAGGTTTGTGATTTCTTGATATGCGAATGCAATGAGGATTTCTTTGTTT 343  
DB 383 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 442  
QY 344 GTGTTGATCTACGATACATTCCTCGAATGATGATGATGATGATGATGATGATGATGAT 403  
DB 443 WTT 502  
QY 404 T 404  
DB 503 W 503

RESULT 10  
CNS005F3/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CNS005F3  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR1P07 of RPc1-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL059925.1 GI:4943047  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

COMMENT

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 960)  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see http://www.fruitfly.org The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
Aron Mammoser in Piet de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPc1-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain Y2; cn bw ep, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.

FEATURES  
source

1..960  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR1P07"  
/clone\_lib="RPc1-98"  
/note="end : TET3"

ORIGIN

Query Match 12.6%; Score 52.6; DB 29; Length 960;  
Best Local Similarity 40.2%; Pred. No. 0.45;  
Matches 158; Conservative 30; Mismatches 205; Indels 0; Gaps 0;  
QY 15 TCATCTCTTTGATTTTCTCGAAATATTTTTCGGTATCGTCAAACTACTGGAATCG 74  
DB 572 TWGARTKTKTGTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 513  
QY 75 CTCGATAGTGGTACGAAATTAGCGAGATTAGTTCTTATCTTCGCCATTATCTGTTT 134  
DB 512 TGWRMTGTGTGRADAAATTTKTGAWDGGTCTTTTGTGTTTGTGTAARKTGRTGTG 453  
QY 135 CTTCCGCGAATGATCTCCGTATAAGATTAGGTAGAGATGAATCGTATAGCTAGAT 194  
DB 452 KGRKTKGTGRDGTGTTTCTTTGTTGTAATTTDTKTGADWWTTTGTGKKAAGRTTTT 393  
QY 195 TTTATCACCAGATAGTTTCTTTGTTCTAGAAATCTCGAAATTTCTCGATAGTTTTCATCT 254  
DB 392 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 333  
QY 255 GTAATAGATTGTTCTTATTCGCGAATGTTGATTAGGTTTGTATTTCTTGTATAGC 314  
DB 332 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 273  
QY 315 GATTGCAATTAGGATTTTCTTTGTTTGTGTTGATCTTACGATACATCTCGCAATG 374  
DB 272 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 213  
QY 375 AATACGTATGGATCTAAATCTGTTAAATTTGTT 407  
DB 212 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 180

RESULT 11  
CNS0081L/c  
LOCUS  
DEFINITION

CNS0081L  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR16N08 of RPc1-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.



ACCESSION AL051208  
 VERSION AL051208.1 GI:4933161  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster (fruit fly)  
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

## REFERENCE

AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
 COMMENT - Web : www.genoscope.cns.fr  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammeter at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

Location/Qualifiers  
 1..832  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACR16N08"  
 /note="end : RPCI-98"  
 /note="end : TET3"

## ORIGIN

Query Match 12.5%; Score 52.4; DB 29; Length 832;  
 Best Local Similarity 39.4%; Pred. No. 0.51;  
 Matches 86; Conservative 38; Mismatches 94; Indels 0; Gaps 0;  
 QY 190 TAGATTTCATCCAGATAGTCTTCTTGTCTAGATCTCTGAATCTCGATAGTTTCA 249  
 DB 278 TACKTWTATATWTATATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 219  
 QY 250 CATGTGTAATAGATGTTCTTATTTCGGGATGTTGATTAGGTTTTCGATTTCTGAT 309  
 DB 218 TTATTCATTTXWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 159  
 QY 310 TAGCGAATTCGAATGAGGATTTCTTTGGTTTGTGTTGATCTTACGATATTCCTGC 369  
 DB 158 TTTTATATATWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 99  
 QY 370 AATTGAATGATGATGATCTAAATCTGTAATTTGTT 407  
 DB 98 TTKTAWCYKGTTCATCTAAKATWTWKWTYKXKW 61

## RESULT 12

EX416727  
 LOCUS BX416727 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone  
 DEFINITION CS0DA011Y114 5-PRIME, mRNA sequence.

ACCESSION BX416727  
 VERSION BX416727.1 GI:30765629  
 KEYWORDS EST.

## SOURCE

ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 712)

## AUTHORS

TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DA011BE07QPI.

## FEATURES

Location/Qualifiers  
 1..712  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DA011Y114"  
 /tissue\_type="NEUROBLASTOMA"  
 /notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

## ORIGIN

Query Match 12.4%; Score 51.8; DB 13; Length 712;  
 Best Local Similarity 10.8%; Pred. No. 0.69;  
 Matches 33; Conservative 152; Mismatches 120; Indels 0; Gaps 0;  
 QY 104 TTAGTTCTATTCTTGGCAATATCTTGTCTTTCGCCGAATGATCTTCGGTATAAGAT 163  
 DB 377 TTTTCTTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWTWT 436  
 QY 164 TTATAGTTAGATGAATCGTATAGTAGTATCATCCAGATAGTTCTTTGTCTAGA 223  
 DB 437 KKKTKTKKKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTKTK 496  
 QY 224 ATCTCGAATCTCGATAGTTTTCATGTCGTAATAGATGTTCTTATTCGGCGATTG 293  
 DB 497 KKK 556  
 QY 284 TTGATTAGGTTTGTGATTTCTTGTATTGATTCGATTCGAATTAGGATTTCTTTGGTTT 343  
 DB 557 KKK 616  
 QY 344 GTGTGATCTTACGATACATTCCTCGCAATGAATAGATGATGATGATGATGATGAT 403  
 DB 617 KKK 676  
 QY 404 TGTTG 408  
 DB 677 KKKKK 681

## RESULT 13

CNS0026Z  
 LOCUS CNS0026Z/c 1101 bp DNA linear GSS 26-JUL-1999  
 DEFINITION Drosophila melanogaster. Genome survey sequence P7 end of BAC BACN01A10 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

## ACCESSION

VERSION AL097301  
 SOURCE GSS.  
 ORGANISM Drosophila melanogaster (fruit fly)

## KEYWORDS

ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 1101)

## AUTHORS

Direct Submission  
 TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
 JOURNAL

- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC and sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - this Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.

| FEATURES | Location/Qualifiers   |
|----------|---|
| 1. 894   | <pre> /organism="Drosophila" /mol_type="genomic" /db_xref="taxon:722" /clone="BACN13B16" /clone_lib="DrosBAC" /plasmid="pBelobAC1" /name="end: T7" </pre> |

Query Match 12.0%; Score 50.2; DB 29; Length 894;  
Best Local Similarity 38.1%; Pred. No. 1.4;  
Matches 140; Conservative 39; Mismatches 188; Indels 0; Gaps 0;

|    |     |  |     |
|----|-----|--|-----|
| Qy | 43  | TTTTTTCCGGTGATCGTGAAACTACTCGTCCGATAGGTGTCAGBAATAGGCGAG         | 102 |
| Db | 550 | TTTTTTCTGKTTTTTTTTTAGTCGGTKGTGCCAACTATKYTKWTGTTTGTKAAA         | 491 |
| Qy | 103 | ATTAGTTTTCTATTCTTGGCCATTATCTGTGTTCTTCCGCCAAATGATCTTCGGTATAAAGA | 162 |
| Db | 490 | TCCTTTTWTGATGAGTTCCTTGATGCGGTTTATTTTGTGTTTGTTKTAKWGT           | 431 |
| Qy | 163 | TTTTTAGGTTAGAGATGAATCGTATAGCTAGATTTTCATCACAGATAGTCTCTTTGTCTAG  | 222 |
| Db | 430 | TWTTTTTKTTTTKAATCYKWWTTVKTKTATTTTTTTTTTKCTWKKTKTTTTTCACAGYTT   | 371 |
| Qy | 223 | AATCTCTGAATTTCTCGATAGTTTTACATGTGTAAATAGATTCGTTCTTATTCGGCGGATT  | 282 |
| Db | 370 | TTGTTTTGTGTTTTTTTKTTTTTTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT  | 311 |
| Qy | 283 | GTTCGATTAGGTTTTTGATTTCTTTGATTATGCGAATGCAATTAGGATTTTTCTTTGGTTTT | 342 |
| Db | 310 | TTTTTTTTTTTGTGTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKTTT    | 251 |
| Qy | 343 | TGTTGTGATCTTACGATACATTCCTGCAATTGAATACGTATGGAATCTAAATCTTGTTAAAT | 402 |
| Db | 250 | TBTXYKGTITTTKTTTGTGWKTTTTVTTTTWTKATWAKTKAYGTWTTTTTKTWTKTWTWT   | 191 |
| Qy | 403 | TTGTGTGA   | 409 |
| Db | 190 | TATTTKW  | 184 |

RESULT 15  
CNS014RT/c  
LOCUS  
DEFINITION  
CNS014RT 975 bp DNA linear GSS 26-JUL-1999  
Drosophila melanogaster genome survey sequence T7 end of BAC  
HAC12L12 of DrosBAC library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
ACCESSION  
VERSION AL104531  
KEYWORDS AL104531.1 GI:5616545  
SOURCE GSS.  
ORGANISM Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1. (bases 1 to 975)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
Web : www.genoscope.cns.fr)





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 01:39:08 ; Search time 2448.33 Seconds  
(without alignments)  
6025.290 Million cell updates/sec

Title: US-09-000-062-7

Perfect score: 494

Sequence: 1 ctccagggaagaacaggat.....cgcatcccggaatctgcg 494

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_hcc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hcc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rod:\*

26: em\_gss\_pig:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 266.4 | 53.9        | 300    | 9     | AV551146           |
| 2          | 202.8 | 41.1        | 320    | 14    | Z34742 ATTS3502 St |
| 3          | 51.4  | 12.4        | 712    | 13    | BX416727 BX416727  |
| 4          | 54.4  | 11.0        | 1101   | 29    | CNS0120N           |

|    |      |      |      |    |          |
|----|------|------|------|----|----------|
| 5  | 54   | 10.9 | 576  | 29 | CNS035N7 |
| 6  | 52.4 | 10.6 | 1131 | 14 | CD050625 |
| 7  | 51.8 | 10.5 | 683  | 29 | CNS0028F |
| 8  | 51.8 | 10.5 | 942  | 29 | CNS018GS |
| 9  | 51.4 | 10.4 | 873  | 13 | BUS34806 |
| 10 | 51.2 | 10.4 | 750  | 29 | CNS011ID |
| 11 | 51   | 10.3 | 1246 | 29 | CG744146 |
| 12 | 50   | 10.1 | 741  | 29 | CNS007YN |
| 13 | 49.6 | 10.0 | 563  | 29 | CNS007UZ |
| 14 | 49.6 | 10.0 | 737  | 29 | CNS008BU |
| 15 | 49.4 | 10.0 | 1403 | 29 | CG749401 |
| 16 | 49.4 | 10.0 | 1388 | 29 | CG756681 |
| 17 | 49.2 | 10.0 | 1022 | 13 | BUS29542 |
| 18 | 49   | 9.9  | 1108 | 29 | CNS00710 |
| 19 | 49   | 9.9  | 1201 | 9  | AL514421 |
| 20 | 48.8 | 9.9  | 848  | 29 | CNS02CV8 |
| 21 | 48.8 | 9.9  | 848  | 12 | BG809697 |
| 22 | 48.8 | 9.9  | 1761 | 28 | CC188336 |
| 23 | 48.6 | 9.8  | 1103 | 28 | BZ551529 |
| 24 | 48.6 | 9.8  | 1201 | 13 | BX336467 |
| 25 | 48.4 | 9.8  | 884  | 29 | CNS0605G |
| 26 | 48.4 | 9.8  | 1058 | 14 | CD048666 |
| 27 | 48.4 | 9.8  | 1300 | 12 | BM468018 |
| 28 | 48.2 | 9.8  | 878  | 29 | CNS0028X |
| 29 | 48.2 | 9.8  | 1059 | 29 | CNS0022B |
| 30 | 48.2 | 9.8  | 1376 | 29 | CG747831 |
| 31 | 47.8 | 9.7  | 922  | 29 | CNS0073W |
| 32 | 47.8 | 9.7  | 1180 | 13 | BX436369 |
| 33 | 47.8 | 9.7  | 1215 | 12 | BM222560 |
| 34 | 47.6 | 9.6  | 446  | 12 | BP516575 |
| 35 | 47.6 | 9.6  | 914  | 13 | BUS63375 |
| 36 | 47.6 | 9.6  | 967  | 29 | CNS0772W |
| 37 | 47.4 | 9.5  | 614  | 29 | CNS0152H |
| 38 | 47   | 9.5  | 867  | 29 | CNS0054A |
| 39 | 47   | 9.5  | 1053 | 13 | BUS08694 |
| 40 | 47   | 9.5  | 1289 | 29 | CG748984 |
| 41 | 47   | 9.5  | 1345 | 29 | CG746697 |
| 42 | 46.8 | 9.5  | 767  | 29 | CNS00AQX |
| 43 | 46.8 | 9.5  | 854  | 29 | CNS012CM |
| 44 | 46.8 | 9.5  | 966  | 14 | CD386802 |
| 45 | 46.8 | 9.5  | 995  | 14 | CD385043 |

## ALIGNMENTS

AV551146 300 bp mRNA linear EST 06-SEP-2000  
AV551146 Arabidopsis thaliana roots Columbia Arabidopsis thaliana  
CDNA clone RZ121h02R 5', mRNA sequence.

AV551146 GI:8722559

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

DNA Res. 7 (3), 175-180 (2000)

20363093

10907847

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/

Location/Qualifiers

```
source 1. 300
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="R2121h02R"
/tissue_type="roots"
/clone_lib="Arabidopsis thaliana roots Columbia"
/notes="vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

Query Match 53.9%; Score 266.4; DB 9; Length 300;
Best Local Similarity 99.6%; Pred. No. 4.5e-45;
Matches 267; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN
QY 1 CTCAGGCGAAGACAGATGATTTGTTCTTAATTAGATCAGGGGTTTAGTCTTTCCAT 60
DB 33 CTCAGGCGAAGACAGATGATTTGTTCTTAATTAGATCAGGGGTTTAGTCTTTCCAT 92
QY 61 TACTTTTTTAATGTTTTTCTGTTACTCTCTCCGGATCTGATTTTACGACAATAGAGTTT 120
DB 93 TACTTTTTTAATGTTTTTCTGTTACTCTCTCCGGATCTGATTTTACGACAATAGAGTTT 152
QY 121 CGGGTTTTCCTCCATTCAGTTTGAATAAAGCTCGTCTTTTAAAGTTTTCGTCGATCGA 180
DB 153 CGGGTTTTCCTCCATTCAGTTTGAATAAAGCTCGTCTTTTAAAGTTTTCGTCGATCGA 212
QY 181 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTTCTTCATCGTTTTT 240
DB 213 TAAACCTGTGAAGATTGAGTCTAGTCGATTTATTGGATGATCCATTTCTTCATCGTTTTT 272
QY 241 TCTTGCTTCGAAGTTCTGTATAACAGA 288
DB 273 TCTTGCTTCGAAGTTCTGTATAACAGA 300

RESULT 2
Z34742
LOCUS ATTS3502 Strasbourg-A Arabidopsis thaliana cDNA clone Fail85 5',
DEFINITION mRNA sequence.
ACCESSION Z34742
VERSION Z34742.1 GI:507089
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 320)
CNRS.
The Arabidopsis thaliana transcribed genome: the GDR cDNA program
Unpublished (1996)
Contact: Philippe G., Gigot C.
Gigot Claude / L512
Laboratoire de Biologie Moléculaire des Plantes - CNRS
12 Rue du General Zimmer, 67084 Strasbourg Cedex, France
Email: ARABANK@MBOC.U-STRASBG.FR.
Location/Qualifiers
1. 320
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="ecotype Columbia"
/db_xref="taxon:3702"
/clone="FAI185"
/clone_lib="Strasbourg-A"
/notes="vector: Lambda ZAPII; tissue-sliced leaves of
A.thaliana ecotype Columbia; clone library=Strasbourg-A;
Cloning vector: Lambda ZAPII; Physiological condition:
leaves strips incubated 2/3/4 days in liquid culture
medium."

ORIGIN
```

```
Query Match 41.1%; Score 202.8; DB 14; Length 320;
Best Local Similarity 85.3%; Pred. No. 5.5e-32;
Matches 273; Conservative 0; Mismatches 42; Indels 5; Gaps 4;

QY 17 GTATGATTTGTTTCTTAATTAGATCAGGGGTTTAGTCTTTCCATTTACTTTTAAAG-TTT 75
DB 1 GTATGATTTGTTTCTTAATTAGATCAGGGGTTTAGTCTTTTCAATTTTAAAGTTT 60
QY 76 TTTCTGTTTACTGTCTCC--GCGATCTCTGATTTTACGACAATAGAGTTTCGGGTTTGTCCC 133
DB 61 TTTCTGTTTACTGTCTCCCGCGATCTCTGATTTTACGCAATAGAGTTTCGGGTTTGTCCA 120
QY 134 ATTCAGTTTGAATAAATACGT-CGGTCTTTTAACTTGTCTGATCGATAAACCTGTGAA 192
DB 121 TTCCAAGTTTGAATAAAGGTCGCGTCTTTTAAAGTTTCTGCGATCGATAAACCTGTGAA 180
QY 193 GATTGAGTCTAGTCTGATTTATTGGATGATCCATTTCTTCATCGTTTTTTTCTTGTTCGAA 252
DB 181 GATTGAGTCTAGTCTGATTTATTGGATGATCCATTTCTTCATCGTTTTTTTCTTGTTCGAA 239
QY 253 GTTCTGTTTAAACAGATTTGTTCTGTGCGGATTTGTCATTACTAGCCGTGTATCGAGAAC 312
DB 240 GTTCTGTTTAAACAGATTTGTCCTGTGTCGATTCGATTCGATTCGATTCGATTCGAGACC 299
QY 313 TAGGGTTTTCGAGTCAATTT 332
DB 300 TAGGGTTTTCGAGTCAATTT 319

RESULT 3
BX416727
LOCUS 712 bp mRNA linear EST 15-MAY-2003
DEFINITION BX416727 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
CS0DA011Y114 5-PRIME, mRNA sequence.
ACCESSION BX416727
VERSION BX416727.1 GI:30765629
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 712)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DA011BE07QP1.
Location/Qualifiers
1. 712
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DA011Y114"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/notes="vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

FEATURES
source
1. 712
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DA011Y114"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA"
/notes="vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match 12.4%; Score 61.4; DB 13; Length 712;
Best Local Similarity 13.4%; Pred. No. 0.0058;
Matches 53; Conservative 179; Mismatches 165; Indels 0; Gaps 0;
```



/organism="Tetraodon nigroviridis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:99883"  
/clone="214A06"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C0AG214BA03SP1-end :  
FUC-Or1"

## ORIGIN

```

Query Match      10.9%; Score 54; DB 29; Length 576;
Best Local Similarity 40.8%; Pred. No. 0.21;
Matches 168; Conservative 27; Mismatches 217; Indels 0; Gaps 0;

QY 18 TATGATTTGTTGTAATAGATCAGGGGTTAGGCTTTTCCATTACTTTTAAAGTTTTT 77
Db 121 :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      TTTTATTTATTTTWTATTTTATTAAGAGTATTTTATTTTATTTTATTTTATTTTATTTT 180

QY 78 TCTGTTACTGCTCGCGATCTGATTTTACGACATAGAGTTTCCGGTTTTTGTCCCAATC 137
Db 181 :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 240

QY 138 CAGTTTGAATAAAGCGTCGCTTTTAAAGTTTGTGATCGATAAACCTGTGAAGATTG 197
Db 241 :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      TTTTATTTTAAAGTTTATTTTAAAGTTTATTTTAAAGTTTATTTTAAAGTTTATTTT 300

QY 198 AGTCTAGTCGATTTATTCGATGATCCATCTTCATCGTTTTTTTCTGCTCGAAGTTCT 257
Db 301 :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      TTTTATTTTAAAGTTTATTTTAAAGTTTATTTTAAAGTTTATTTTAAAGTTTATTTT 360

QY 258 GTATAACAGAGTTTCTGCTGTGCGATTTGTCATTTACCTAGCGGTGTATCGAAGATTG 317
Db 361 :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      TTTTATTTTAAAGTTTATTTTAAAGTTTATTTTAAAGTTTATTTTAAAGTTTATTTT 420

QY 318 TTTTCGAGTCATTTTGCCTTTTGGTTATATCTGTTTCGATAACGATTCATCTGGATT 377
Db 421 :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      TTTTATTTTAAAGTTTATTTTAAAGTTTATTTTAAAGTTTATTTTAAAGTTTATTTT 480

QY 378 AGGGTTTAAAGTGTGAGCTTTAGTATCCAAATTTCTTCAAAATTTAGTTAT 429
Db 481 TTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 532

```

## RESULT 6

```

CD050625/c
LOCUS      1131 bp mRNA linear EST 09-MAY-2003
DEFINITION AGENCOURT_1397661 NIH_MGC_173 Homo sapiens cDNA 5', mRNA sequence.
ACCESSION  CD050625
VERSION     CD050625.1 GI:30487745
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1131)
            NIH-MGC http://mgi.nci.nih.gov/
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dr. Jamie Thompson, University of WI
            cDNA Library Preparation: Gina Zastrow-Hayes
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDKM32 row: m column: 05
            High quality sequence start: 66
            High quality sequence stop: 257.

```

## FEATURES

source  
1..1131  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="taxon:9606"  
/tissue\_type="embryonic trophoblasts, made from WA01 stem  
cells"  
/lab\_host="DH10B Tona"  
/clone\_lib="NIH MGC 173"  
/note="Vector: pDONR201; Site 1: attP2; Site 2: attP1;  
LIBR PRIMING - oligo dT; METHOD - full-length enriched;  
LIBR PROVIDER - Bradfield"

## ORIGIN

```

Query Match      10.6%; Score 52.4; DB 14; Length 1131;
Best Local Similarity 45.4%; Pred. No. 0.37;
Matches 188; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 23 TTGTTGTAATAGATCAGGGGTTAGGCTTTTCCATTACTTTTAAAGTTTTTCTGT 82
Db 700 TTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 641

QY 83 TACTGCTCCGCGATCTGATTTTACGACATAGAGTTTCCGGTTTTTGTCCCAATCCAGTT 142
Db 640 TTTTATTTTATTTTGGTGTGTTTGTGTTTATTTTGTGTTTATTTTATTTTATTTT 581

QY 143 TGAATAAATACGTCCTCTTTTAAAGTTTCTCGATCGATAAACCTGTGAAGATTGAGTCT 202
Db 580 TTTTATTTTATTTTGGTGTGTTTGTGTTTATTTTGTGTTTATTTTATTTTATTTT 521

QY 203 AGTCGATTTATTCGATGATCCATCTTCATCTGTTTCTGCTTCGAGTCTGTATA 262
Db 520 GGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 461

QY 263 ACCAGATTTGTCGTGCGATTTGTCATTTACCTAGCGGTGATCGAGAACTAGGGTTTC 322
Db 460 TTTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 401

QY 323 GAGTCAATTTGCCCCCTTTGTTATATCTGTTTCGATAACGATTCATCTGATAGGTT 382
Db 400 TGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 341

QY 383 TTTAAGTGTGAGCTTTAGTATTTCCAAATTTCTTCAAAATTTAGTTATGATAAT 436
Db 340 TTTTATTTTATTTATTTAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 287

```

## RESULT 7

```

CNS0028F/c
LOCUS      683 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN01G02 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL097353
VERSION     AL097353.1 GI:5608964
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Aphididae; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 683)
            Genoscope.
            Direct Submission
            Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila genome project (EDGP) -
            http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES    Location/Qualifiers
            source
            1..683

```



```

FEATURES             source
Location/Qualifiers
1..750
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN06D21"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : T7"

ORIGIN
Query Match          10.4%; Score 51.2; DB 29; Length 750;
Best Local Similarity 33.2%; Pred. No. 0.72;
Matches 113; Conservative 62; Mismatches 165; Indels 0; Gaps 0;

QY 18 TATGATTGTTTGTGAATTAGATCAGGGGTTTAGGTCCTTCATTACTATTTTAAAGTGTTTT 77
DB 498 TKKTWTKTTTTTTTAAWTTTWTGKTWTGTTTTTTTWWKWAATTKTTTTTAAGWTTATK 439
QY 78 TCTGTTACTGTCCTCGGATCTGATTTTACGACAATAGAGTTTCGGGTTTGTGCCATTC 137
DB 438 TTTGKTTTTTYYKTTTTTTTTTTCATKTTGYTTTTTGTGATTTTCTTTTTCATTTT 379
QY 138 CAGTTTGAATAAATACGTCGGTCCTTTTAAAGTTTGTGGATCGATAAACCTGTGAAGATTG 197
DB 378 TWKTTTTTTTWTCTAAGTTTKTYTAGTTTTCTYATTTCYTTTAAATKWAKTTTTTTT 319
QY 198 AGTCAGTCGATTATTGGATGATCAATCTTCATCGGTTTTTTTCTGCTCGAAGTTCT 257
DB 318 KTAGYTTTWTGTGTTTTTKTTTTKWTCTGAKTKWKTTTTTTCWTTTTTWWKTAWTTTK 259
QY 258 GTATAAACCAGATTTGTCTGTGTCGATTGTCATTACCTAGCCGTATCGAGAACTAGGG 317
DB 258 KYTTAAWAKCTTTTKTTTTKKWTKTTTTTTTATTAKTWTTCCTKTWTKKWTTTTT 199
QY 318 TTTTCGAGTCATTTGCCCTTTTGGTTTATATCTGGTTC 357
DB 198 KTTTTCGAKTCTCTTTTWTYTTTKGTWTKYKTTTGTY 159

RESULT 11
CG744146
LOCUS
DEFINITION
P036-3-E10.ya Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
CG744146
CG744146.1 GI:37965014
GSS.
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 1246)
Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode Pristionchus
pacificus
Mol. Genet. Genomics 269 (5), 715-722 (2003)
22835951
12884007
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@uebingen.mpg.de
Class: BAC ends
Location/Qualifiers
1..1246
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"

FEATURES             source
Location/Qualifiers

```



```

/db_xref="taxon:54125"
/clone_lib="Ppa EcoRI BAC library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

ORIGIN
Query Match      10.3%; Score 51; DB 29; Length 1246;
Best Local Similarity 39.2%; Pred. No. 0.69;
Matches 177; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

Qy 18 TATGATTGTTTGTAAATAGATCAGGGGTTTAGGTCCTTCCATTACTTTTTAAATGTTTT 77
Db 742 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 801

Qy 78 TCTGTACTGTCCTCGGATCTGATTTTACGACATAGATTCGGGTTTGTCCCATTC 137
Db 802 TTTTTTTTNNNTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 861

Qy 138 CAGTTTGAATAAAGCGTCCTTTTAAAGTTTGTGATCGATAAAGCTGTGAAGATTG 197
Db 862 TNNNNNNNTTNTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 921

Qy 198 AGCTAGTCGATTTATGATGATCATCTTTCATCGTTTTTTCTTCGTCGAAGTTCT 257
Db 922 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 981

Qy 258 GTATAACCAAGATGTCCTGCGGATCTGATTCATTACCTAGCGCTGTCGAGAACTAGGG 317
Db 982 NTNTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1041

Qy 318 TTTTCGAGTCAATTTTCGCCCTTTTGGTTATATCTGGTTCGATACGATCACTGGATT 377
Db 1042 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1101

Qy 378 AGGGTTTAAAGTGGTCACTTTAGTATCCAAATTCCTCAAAATTTAGTATGGAATAG 437
Db 1102 TTTTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTN 1161

Qy 438 AAAATCCGAATGACTGTTCATTTCTGTT 469
Db 1162 NTNTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1193

CNS007YN 741 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR16L07 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL051001 GI:4933055
VERSION AL051001.1
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 741)
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osagawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial

```

EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain v2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES
Location/Qualifiers
1..741
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR16L07"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN
Query Match      10.1%; Score 50; DB 29; Length 741;
Best Local Similarity 34.6%; Pred. No. 1.3;
Matches 155; Conservative 0; Mismatches 293; Indels 0; Gaps 0;

Qy 18 TATGATTGTTTGTAAATAGATCAGGGGTTTAGTCTTCCATTACTTTTTAAATGTTTT 77
Db 138 TTTTNNTTTNTNTNTNTNNNTTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 197

Qy 78 TCTGTACTGTCCTCGGATCTGATTTTACGACATAGATTCGGGTTTGTCCCATTC 137
Db 198 TTTTNNNNNTTNTNTNTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTT 257

Qy 138 CAGTTTGAATAAAGCTCGCTCTTTTAAAGTTTGTGATCGATAAAGCTGTGAAGATTG 197
Db 258 NNNNTTTTTTNTTNTNTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNCN 317

Qy 198 AGCTAGTCGATTTATGATGATCATCTTTCATCGTTTTTTCTTCGTCGAAGTTCT 257
Db 318 NNNNTNTNTNTNTNTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNTNTT 377

Qy 258 GTATAACCAAGATGTCCTGCGGATCTGATTCATTACCTAGCGCTGTCGAGAACTAGGG 317
Db 378 NTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTCNCNTNNNTTTTTTTT 437

Qy 318 TTTTCGAGTCAATTTTTCGCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCGATT 377
Db 438 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNNNNNNNTTTTTTTTTTTT 497

Qy 378 AGGGTTTAAAGTGGTCACTTTAGTATCCAAATTCCTCAAAATTTAGTATGGAATAG 437
Db 498 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTNTNTNTTTTTTTTTTNT 557

Qy 438 AAAATCCGAATGACTGTTCATTTCT 465
Db 558 TNNTTNNNTNTNTNTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 585

```

```

RESULT 13
CNS007YN 563 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR16P14 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL050869
VERSION AL050869.1 GI:4932923
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 563)
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
REFERENCE
AUTHORS
TITLE
JOURNAL

```



COMMENT  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

# FEATURES

Location/Qualifiers  
1..563  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR16P14"  
/clone\_lib="RPCL-98"  
/note="end : TET3"

## ORIGIN

Query Match 10.0%; Score 49.6; DB 29; Length 563;  
Best Local Similarity 35.9%; Pred. No. 1.7;  
Matches 148; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

Qy 18 TATGATTGTTGTAATAGATCAGGGGTTTAGGCTTTCCATTACTTTTTTAATGTTTT 77  
Db 74 TTTTGTGTTGTTTTTTTTTTTTTTTTTTTTTGTGTTTTTTTTTTTTTTTTTTTT 133  
Qy 78 TCTGTACTGTCGCGGATCGATTCGATTTACGACATAGAGTTTCGGGTTTGTCCCATTC 137  
Db 134 TTTTGTGTTGTTTTTTTTTTTTTTTTTTTTTCKNNNNNNNTNNNNNTNNNTTT 193  
Qy 138 CAGTTGAAATAAACGTCGGCTTTTAAGTTTGTGATCGATAAACCTGTGAAGTTG 197  
Db 194 NTNNNNNNNTNNNNNTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 253  
Qy 198 AGCTAGTCGATTTATGCGATGATCCATTCCTCATCGTTTTTTCTTGCTTCGAAGTTCT 257  
Db 254 TTTTGTGTTGTTTTTTTTTTTTTTTTTTTTTNNNNNTNNNTTTTTTTCTTNNNNNNNTNN 313  
Qy 258 GTATACACGATTTGCTGTGCGATGTCATTAACCTAGCGGTGTCGAGAACTAGGG 317  
Db 314 NTNNATNNNTTT 373  
Qy 318 TTTTCGAGTCAATTTGCCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGATT 377  
Db 374 TTTTGTGTTGTTTTTTTTTTTTTTTTTTTTTNNNNNTNNNTNNNTNNNTNNNTTT 433  
Qy 378 AGGGTTAAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTAGTTAT 429  
Db 434 NTNTNTTTTGTGTCGNNNNNTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 485

## RESULT 14

CNS008BU 737 bp DNA linear GSS 03-JUN-1999  
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #  
DEFINITION BACR16E03 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL051476.1 GI:4933530

VERSION GSS.

KEYWORDS Drosophila melanogaster (fruit fly)

SOURCE Drosophila melanogaster

ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 737)

## REFERENCE

## AUTHORS

TITLE  
JOURNAL

## COMMENT

Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqrefgenoscope.cns.fr](mailto:seqrefgenoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCL-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

Location/Qualifiers  
1..737  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR16E03"  
/clone\_lib="RPCL-98"  
/note="end : TET3"

## ORIGIN

Query Match 10.0%; Score 49.6; DB 29; Length 737;  
Best Local Similarity 33.3%; Pred. No. 1.5;  
Matches 138; Conservative 14; Mismatches 263; Indels 0; Gaps 0;  
Qy 18 TATGATTGTTGTAATAGATCAGGGGTTTAGGCTTTCCATTACTTTTTTAATGTTTT 77  
Db 310 TTTTGTGTTGTTTTTTTTTTTTTTTTTTTTTNNNNNTNNNTNNNTNNNTTT 369  
Qy 78 TCTGTACTGTCGCGGATCTGATTTACGACATAGAGTTTCGGGTTTGTCCCATTC 137  
Db 370 NTNNNTNNNTNNNTNNNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 429  
Qy 138 CAGTTTGAATAAACGTCGGCTTTTAAAGTTTGTGCGATCGATAAACCTGTGAAGATTG 197  
Db 430 TTTTGTGTTGTTTTTTTTTTTTTTTTTTTTTNCANTTGNNTNNNTTTTTTTTT 489  
Qy 198 AGCTAGTCGATTTATGCGATGATCCATTCCTCATCGTTTTTTCTTGCTTCGAAGTTCT 257  
Db 490 TTTTGTGTTGTTTTTTTTTTTTTTTTTTTTTNNNNNTNNNTNNNTNNNTTT 549  
Qy 258 GTATAACACGATTTGCTGTGCGATTCGATTAACCTAGCCGTGTCGAGAACTAGGG 317  
Db 550 TTTTGTGTTGTTTTTTTTTTTTTTTTTTTTTNNNNNTNNNTNNNTNNNTNN 609  
Qy 318 TTTTCGAGTCAATTTGCCCCCTTTTGGTTATATCTGGTTCGATAACGATTCATCTGATT 377  
Db 610 TTTTGTGTTGTTTTTTTTTTTTTTTTTTTTTNNNNNTNNNTNNNTNNNTTT 669  
Qy 378 AGGGTTTAAAGTGGTGACGTTTAGTATTCCAATTTCTTCAAAATTAGTTATGGA 432  
Db 670 NNSTTTNNNNNTNNNTTTTATTTNNNTNNNTNNNTNNNTNNNTNNNTNN 724

## RESULT 15

CG749401

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CG749401 1403 bp DNA linear GSS 24-OCT-2003  
P043-3-E04.2a Ppa EcoRI BAC Library Pristionchus pacificus genomic,  
genomic survey sequence.  
CG749401  
CG749401.1 GI:37970327  
GSS.  
Pristionchus pacificus  
Pristionchus pacificus



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2004, 23:55:13 ; Search time 366.167 Seconds

(without alignments)  
5731.297 Million cell updates/sec

Title: US-09-000-062-7

Perfect score: 494

Sequence: 1 ctcaggcgaagaacaggtat.....cgcagatcccggtatctgcg 494

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001as:\*

5: Geneseq2001bs:\*

6: Geneseq2002s:\*

7: Geneseq2003as:\*

8: Geneseq2003bs:\*

9: Geneseq2003cs:\*

10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description        |
|------------|-------|---------------|--------|----|--------------------|
| 1          | 490.8 | 99.4          | 494    | 2  | Aat85997 Arabidops |
| 2          | 374.2 | 75.7          | 1005   | 3  | Aac47929 Arabidops |
| 3          | 84.6  | 17.1          | 677    | 3  | Aac32694 Arabidops |
| 4          | 83.6  | 16.9          | 674    | 3  | Aac51182 Arabidops |
| 5          | 55.8  | 11.3          | 12870  | 6  | Abk39983 Human che |
| 6          | 55.8  | 11.3          | 12870  | 6  | Abk39983 Human che |
| 7          | 49.9  | 9.9           | 17869  | 6  | Abk39921 Human che |
| 8          | 49.9  | 9.9           | 17869  | 6  | Abk39921 Human che |
| 9          | 48.9  | 9.7           | 6593   | 6  | Abk32478 Human inn |
| 10         | 46.6  | 9.4           | 7814   | 4  | Aas46530 Tumour su |
| 11         | 46.4  | 9.4           | 6650   | 6  | Abk32638 Human inn |
| 12         | 46.2  | 9.4           | 6048   | 6  | Abk32509 Human inn |
| 13         | 46.9  | 9.3           | 3025   | 7  | Abk32105 Human inn |
| 14         | 46.9  | 9.3           | 3025   | 9  | Abk32105 Human inn |
| 15         | 46.9  | 9.3           | 3025   | 9  | Abk32105 Human inn |
| 16         | 45.6  | 9.2           | 6688   | 6  | Abk32638 Human inn |
| 17         | 45.6  | 9.2           | 7442   | 4  | Abk32638 Human inn |
| 18         | 44.8  | 9.1           | 5208   | 6  | Abk32091 Human inn |
| 19         | 44.8  | 9.1           | 12733  | 6  | Abk32091 Human inn |
| 20         | 44.4  | 9.0           | 9547   | 6  | Abk33505 Human inn |
| 21         | 44.2  | 8.9           | 594    | 6  | Abq46990 Oligonuc  |
| 22         | 44.2  | 8.9           | 594    | 6  | Abq46990 Oligonuc  |
| 23         | 44.2  | 8.9           | 5204   | 6  | Abk32899 Human inn |

## ALIGNMENTS

### RESULT 1

AAT85997

ID AAT85997 standard; DNA; 494 BP.

XX XX

AC AAT85997;

XX XX

DT 17-NOV-1997 (first entry)

XX XX

DE Arabidopsis thaliana histone H3.3-like DNA fragment (intron 2).

XX XX

KW Plant expression regulation sequence; intron 2; histone;  
KW herbicide tolerance; 5-enolpyruvylshikimate-3-phosphate synthase; EPSPS;  
KW glyphosate; ds.

XX XX

OS Arabidopsis thaliana.

XX XX

FN WO9704114-A2.

XX XX

PD 06-FEB-1997.

XX XX

EF 17-JUL-1996; 96WO-FR001109.

XX XX

PR 19-JUL-1995; 95FR-00008980.

XX XX

(RHON ) RHONE POULENC AGROCHIMIE.

PA PA

XX XX

PI Derose R, Chaubet N, Gigot C;

XX XX

DR WPI; 1997-132652/12.

XX XX

PT New regulatory sequence for chimeric gene expression in rapidly growing parts of a plant - includes at least one intron from a plant histone gene and is useful for imparting resistance to herbicides.

PT PT

XX XX

PS Claim 4; Page 27; 3lpp; French.

XX XX

The known cosmid clone c22 of Arabidopsis thaliana contains two histone H3.3-like genes. Digestion of clone c22 with restriction enzymes AluI and CfoI generated a fragment of 494 bp having the present sequence. This fragment, designated intron 2, was ligated to synthetic linkers for cloning into plant expression vectors. In addition to the intron 2 sequence, the vectors contained a plant promoter and a herbicide tolerance gene (e.g. a mutated version of the 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) gene). The intron enhances expression of the herbicide tolerance gene in rapidly growing parts of plants. The intron can also be used to enhance expression of genes that impart resistance to

```
CC pathogens or that encode nutritional or therapeutic proteins
XX SQ Sequence 494 BP; 113 A; 83 C; 104 G; 194 T; 0 U; 0 Other;
    Query Match          99.4%; Score 490.8; DB 2; Length 494;
    Best Local Similarity 99.6%; Pred. No. 5.5e-118;
    Matches 492; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGCGAAGAACAGGTATGTTTGTGTTGAATAGATCAGGGGTTAGTCTCTTTCAT 60
DB 1 CTCAGCGAAGAACAGGTATGTTTGTGTTGAATAGATCAGGGGTTAGTCTCTTTCAT 60

QY 61 TACTTTTTTAAGTCTTTTCTGTTACTGCTCCGCGATCTGATTTTACGACAAATAGAGTTT 120
DB 61 TACTTTTTTAAGTCTTTTCTGTTACTGCTCCGCGATCTGATTTTACGACAAATAGAGTTT 120

QY 121 CCGGTTTTGTCCTCCATCCAGTTTGAATAAATTAACGTCCTTTTAAAGTTTGCCTGATCGA 180
DB 121 CCGGTTTTGTCCTCCATCCAGTTTGAATAAATTAACGTCCTTTTAAAGTTTGCCTGATCGA 180

QY 181 TAAACCTGTGAAGATGAGTCTAGTCGATTTTATTTGGATGATCCATTTCTTCATCGTTTTT 240
DB 181 TAAACCTGTGAAGATGAGTCTAGTCGATTTTATTTGGATGATCCATTTCTTCATCGTTTTT 240

QY 241 TCTTGCTTCGAAAGTCTGATTAACAGATTTGTCGTCGTCGATTTGATACCTGATCGCG 300
DB 241 TCTTGCTTCGAAAGTCTGATTAACAGATTTGTCGTCGTCGATTTGATACCTGATCGCG 300

QY 301 TGTATCGAAGTACGAGTTTTCGAGTCAATTTTGGCCCTTTTGGTTATATCTGTTTCGAT 360
DB 301 TGTATCGAAGTACGAGTTTTCGAGTCAATTTTGGCCCTTTTGGTTATATCTGTTTCGAT 360

QY 361 AACGATTCATCTGATGATAGGTTTAAAGTGGTGAGCTTTAGTATTCCTTCAATTTCTTCAAAA 420
DB 361 AACGATTCATCTGATGATAGGTTTAAAGTGGTGAGCTTTAGTATTCCTTCAATTTCTTCAAAA 420

QY 421 TTTAGTTATGATTAATGAATAATCCGAAATGATCTTCAATTTCTGTTAAATCGGCAGA 480
DB 421 TTTAGTTATGATTAATGAATAATCCGAAATGATCTTCAATTTCTGTTAAATCGGCAGA 480

QY 481 TCCCGGATCTCGG 494
DB 481 TCCCGGATCTCGG 494

RESULT 2
AAC47929
ID AAC47929 standard; DNA; 1005 BP.
XX
XX AAC47929;
AC
XX
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 55633.
XX
XX Hybridisation assay; Genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-012825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX
```

Hybridisation assay: genetic mapping: gene expression control:

KW protein identification; signal transduction pathway; metabolic pathway;  
KW promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 99US-0123180P.

XX 09-MAR-1999; 99US-0123548P.

XX 23-MAR-1999; 99US-0125789P.

XX 25-MAR-1999; 99US-0126264P.

XX 29-MAR-1999; 99US-0126785P.

XX 01-APR-1999; 99US-0127462P.

XX 06-APR-1999; 99US-0128234P.

XX 08-APR-1999; 99US-0128714P.

XX 16-APR-1999; 99US-0129845P.

XX 19-APR-1999; 99US-0130077P.

XX 21-APR-1999; 99US-0130449P.

XX 23-APR-1999; 99US-0130510P.

XX 23-APR-1999; 99US-0130891P.

XX 28-APR-1999; 99US-0131449P.

XX 30-APR-1999; 99US-0132048P.

XX 30-APR-1999; 99US-0132407P.

XX 04-MAY-1999; 99US-0132484P.

XX 05-MAY-1999; 99US-0132485P.

XX 06-MAY-1999; 99US-0132486P.

XX 06-MAY-1999; 99US-0132487P.

XX 07-MAY-1999; 99US-0132863P.

XX 11-MAY-1999; 99US-0134256P.

XX 14-MAY-1999; 99US-0134218P.

XX 14-MAY-1999; 99US-0134219P.

XX 14-MAY-1999; 99US-0134221P.

XX 18-MAY-1999; 99US-0134768P.

XX 19-MAY-1999; 99US-0134941P.

XX 20-MAY-1999; 99US-0135124P.

XX 21-MAY-1999; 99US-0135353P.

XX 24-MAY-1999; 99US-0135629P.

XX 25-MAY-1999; 99US-0136021P.

XX 27-MAY-1999; 99US-0136392P.

XX 28-MAY-1999; 99US-0136782P.

XX 01-JUN-1999; 99US-0137222P.

PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.

|    |              |                |
|----|--------------|----------------|
| PR | 07-SEP-1999; | 90US-0152363P; |
| PR | 10-SEP-1999; | 90US-0153070P; |
| PR | 13-SEP-1999; | 90US-0153758P; |
| PR | 15-SEP-1999; | 90US-0154018P; |
| PR | 16-SEP-1999; | 90US-0154039P; |
| PR | 20-SEP-1999; | 90US-0154779P; |
| PR | 22-SEP-1999; | 90US-0155139P; |
| PR | 23-SEP-1999; | 90US-0155486P; |
| PR | 24-SEP-1999; | 90US-0155653P; |
| PR | 28-SEP-1999; | 90US-0156458P; |
| PR | 29-SEP-1999; | 90US-0156596P; |
| PR | 04-OCT-1999; | 90US-0157117P; |
| PR | 05-OCT-1999; | 90US-0157753P; |
| PR | 06-OCT-1999; | 90US-0157865P; |
| PR | 07-OCT-1999; | 90US-0158023P; |
| PR | 08-OCT-1999; | 90US-0158233P; |
| PR | 12-OCT-1999; | 90US-0158369P; |
| PR | 13-OCT-1999; | 90US-0159293P; |
| PR | 13-OCT-1999; | 90US-0159294P; |
| PR | 14-OCT-1999; | 90US-0159295P; |
| PR | 14-OCT-1999; | 90US-0159323P; |
| PR | 14-OCT-1999; | 90US-0159330P; |
| PR | 14-OCT-1999; | 90US-0159331P; |
| PR | 14-OCT-1999; | 90US-0159637P; |
| PR | 18-OCT-1999; | 90US-0159638P; |
| PR | 18-OCT-1999; | 90US-0159584P; |
| PR | 21-OCT-1999; | 90US-0160741P; |
| PR | 21-OCT-1999; | 90US-0160767P; |
| PR | 21-OCT-1999; | 90US-0160770P; |
| PR | 21-OCT-1999; | 90US-0160811P; |
| PR | 21-OCT-1999; | 90US-0160815P; |
| PR | 22-OCT-1999; | 90US-0160980P; |
| PR | 22-OCT-1999; | 90US-0160981P; |
| PR | 22-OCT-1999; | 90US-0160983P; |
| PR | 25-OCT-1999; | 90US-0161404P; |
| PR | 25-OCT-1999; | 90US-0161405P; |
| PR | 25-OCT-1999; | 90US-0161406P; |
| PR | 26-OCT-1999; | 90US-0161359P; |
| PR | 26-OCT-1999; | 90US-0161360P; |
| PR | 26-OCT-1999; | 90US-0161361P; |
| PR | 28-OCT-1999; | 90US-0161920P; |
| PR | 28-OCT-1999; | 90US-0161922P; |
| PR | 28-OCT-1999; | 90US-0161993P; |
| PR | 29-OCT-1999; | 90US-0162142P; |

Query Match 17.1%;

Best Local Similarity 77.9%;

Matches 102; Conservative 0

|    |     |                        |
|----|-----|------------------------|
| Qy | 69  | AATGTTTTCCTGTTACTGTGTT |
| Db | 1   | AATGTTTTCCTGTTACTGTGTT |
| Qy | 129 | GTCCCATTCACGTTTGAAAAA  |
| Db | 61  | GTCCCATTCACGTTTGAAAAA  |
| Qy | 189 | TGAAGATTGAG 199        |
| Db | 121 | CGAGGTAAAG 131         |

```

RESULT 4
AAC51182
ID ID AAC51182 standard; DNA; 674 B
XX
XX
XX AAC51182;
XX AC
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA frag
XX

```

|    |   |
|----|---|
| XX | Hybridisation assay; genetic mapping; gene expression control;          |
| XX | protein identification; signal transduction pathway; metabolic pathway; |
| XX | promoter; termination sequence; ss.                                     |
| XX |   |
| OS | <i>Arabidopsis thaliana</i> .   |
| XX |   |
| PN | EP1033405-A2.   |
| XX |   |
| PD | 06-SEP-2000.  |
| XX |   |
| PF | 25-FEB-2000; 2000EP-00301439.   |
| XX |   |
| PR | 25-FEB-1999; 99US-0121825P.   |
| PR | 05-MAR-1999; 99US-01231180P.  |
| PR | 03-MAR-1999; 99US-0123548P.   |
| PR | 23-MAR-1999; 99US-0125788P.   |
| PR | 25-MAR-1999; 99US-0126264P.   |
| PR | 29-MAR-1999; 99US-0126785P.   |
| PR | 01-APR-1999; 99US-0127462P.   |
| PR | 06-APR-1999; 99US-0128234P.   |
| PR | 08-APR-1999; 99US-0128714P.   |
| PR | 16-APR-1999; 99US-0129845P.   |
| PR | 19-APR-1999; 99US-0130077P.   |
| PR | 21-APR-1999; 99US-0130449P.   |
| PR | 23-APR-1999; 99US-0130510P.   |
| PR | 23-APR-1999; 99US-0130891P.   |
| PR | 28-APR-1999; 99US-0131449P.   |
| PR | 30-APR-1999; 99US-0132048P.   |
| PR | 30-APR-1999; 99US-0132407P.   |
| PR | 04-MAY-1999; 99US-0132484P.   |
| PR | 05-MAY-1999; 99US-0132485P.   |
| PR | 06-MAY-1999; 99US-0132486P.   |
| PR | 06-MAY-1999; 99US-0132487P.   |
| PR | 07-MAY-1999; 99US-0132863P.   |
| PR | 11-MAY-1999; 99US-0134256P.   |
| PR | 14-MAY-1999; 99US-0134218P.   |
| PR | 14-MAY-1999; 99US-0134219P.   |
| PR | 14-MAY-1999; 99US-0134221P.   |
| PR | 18-MAY-1999; 99US-0134370P.   |
| PR | 18-MAY-1999; 99US-0134768P.   |
| PR | 19-MAY-1999; 99US-0134941P.   |
| PR | 20-MAY-1999; 99US-0135124P.   |
| PR | 21-MAY-1999; 99US-0135353P.   |
| PR | 24-MAY-1999; 99US-0135629P.   |
| PR | 25-MAY-1999; 99US-0136021P.   |
| PR | 27-MAY-1999; 99US-0136392P.   |
| PR | 28-MAY-1999; 99US-0136782P.   |
| PR | 01-JUN-1999; 99US-0137232P.   |
| PR | 03-JUN-1999; 99US-0137528P.   |
| PR | 04-JUN-1999; 99US-0137502P.   |
| PR | 07-JUN-1999; 99US-0137724P.   |
| PR | 08-JUN-1999; 99US-0138094P.   |
| PR | 10-JUN-1999; 99US-0138540P.   |
| PR | 10-JUN-1999; 99US-0138847P.   |
| PR | 14-JUN-1999; 99US-0139119P.   |
| PR | 16-JUN-1999; 99US-0139452P.   |
| PR | 16-JUN-1999; 99US-0139453P.   |
| PR | 17-JUN-1999; 99US-0139492P.   |
| PR | 18-JUN-1999; 99US-0139454P.   |
| PR | 18-JUN-1999; 99US-0139455P.   |
| PR | 18-JUN-1999; 99US-0139456P.   |
| PR | 18-JUN-1999; 99US-0139457P.   |
| PR | 18-JUN-1999; 99US-0139458P.   |
| PR | 18-JUN-1999; 99US-0139459P.   |
| PR | 18-JUN-1999; 99US-0139460P.   |
| PR | 18-JUN-1999; 99US-0139461P.   |
| PR | 18-JUN-1999; 99US-0139462P.   |
| PR | 18-JUN-1999; 99US-0139463P.   |
| PR | 18-JUN-1999; 99US-0139750P.   |
| PR | 18-JUN-1999; 99US-0139763P.   |
| PR | 21-JUN-1999; 99US-0139817P.   |
| PR | 22-JUN-1999; 99US-0139899P.   |
| PR | 23-JUN-1999; 99US-0140353P.   |

```
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143254P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145119P.
PR 23-JUL-1999; 99US-0145115P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145313P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147535P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
```

```
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158025P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159328P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
```

Query Match 16.9%; Score 83.6; DB 3; Length 674;

Best Local Similarity 77.7%; Pred. No. 5.6e-12;

Matches 101; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```
Qy 70 ATGTTTTTCTGTACTGTCTCGCGATCTGATTTTACGACATAGAGTTTCGGGTTTG 129
Db 1 ATGTTTTTCTGTACTGTCTCGCGATCTGATTTTACGACATAGAGTTTCGGGTTTG 60
Qy 130 TCCATTCCAGTTTGAAATAAAGCCGTCCTTTTAAAGTTTCTGGATCGATAAACCTGT 189
Db 61 TCCATTCCAGTTTGAAATAAAGATGGCTCGTACCAGCAACCGCTGTAAGTCCACC 120
Qy 190 GAAGATTGAG 199
Db 121 GGAGGTAAG 130
```

#### RESULT 5

ABK39983  
ID ABK39983 standard; DNA; 12870 BP.

XX AC ABK39983;

DT 21-MAY-2002 (first entry)

DE Human chemically pretreated gene sequence #33 strand 1.







KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
KW ds.

OS Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for  
XX diagnosis and treatment of diseases associated with abnormal cytosine  
XX methylation.

XX Claim 1; SEQ ID NO 78; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
XX genes which are modified by the methylation of cytosines. The sequences  
XX can be used in the diagnosis and treatment of immune system disorders,  
XX including eye diseases such as retinopathy, neovascular glaucoma and  
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
XX diseases. The present sequence is a gene of the invention

XX Sequence 17869 BP; 5366 A; 158 C; 3365 G; 8978 T; 0 U; 2 Other;

Query Match 9.9%; Score 49; DB 6; Length 17869;  
Best Local Similarity 46.0%; Pred. No. 0.018;  
Matches 210; Conservative 0; Mismatches 240; Indels 7; Gaps 1;

QY 18 TATGATTGTTGTAATAGATCAGGGGTTTAGGCTTTTCCATTACTTTTAAAGTTT 77  
DB 17051 TTGTTTGTGTTTGTGAAATTTATGATAGTTTATTTTGTGTTT 17110  
QY 78 TCTGTTACTGTCGCCGATCTGATTTTACGACATAGAGTTTCGGTTTGCCTATC 137  
DB 17111 TTGTTTGTGTTTATTTTTCGTTTTTTTCGTTGTCGTTGATTTTGGTTGTATTT 17170  
QY 138 CAGTTTGAATAAAGCTCCGCTTTTAAAGTTTGTGATCGATAAACCTGTGAAGATTG 197  
DB 17171 GTTTTGGTTGTTTTCGTTTGTGTTTGTGTTTGTGTTTATTTAGTTTATGTTT 17230  
QY 198 AGTCTAGTCGATTTATGATGATGATCATCTTCATCGTTTTCCTGCTCGAAGTTCT 257  
DB 17231 TGT-----TTTGTGTTGTTTATTTTGTGTTGTTTGTGTTTGTGTTT 17283  
QY 258 GTATAACAGATTGTCGTGCGGATTCGATTAACCTAGCGTGTATCGACAACATGAG 317  
DB 17284 GGTGTTTGTGTTGTTTATTTTATTTTATTTTATTTTATTTTGTGTTGATGTTAG 17343  
QY 318 TTTTCGAGTCAATTTTGCCTTTTTCGTTTATATCTGTTTCGATACGATTCATCTGATT 377  
DB 17344 TTTTGTGTTTGTGTTTATTTATTTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 17403  
QY 378 AGGGTTTAAAGTGGTACGCTAGTTAGTATCCAAATTTCTTCAAAATTTAGTTATGATA 437  
DB 17404 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 17463  
QY 438 AAAATCCGAATTCAGTTTCATTTCTTGTAAATG 474  
DB 17464 TTTTGTGAGTTTGTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 17500

RESULT 9

ABL32478

XX ABL32478 standard; DNA; 6593 BP.

XX ABL32478;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 451.

XX Human; immune system disease; cytosine methylation; antiaesthmatic;  
XX antiarteriosclerotic; antianaemic; cytosinatic; neotropic;  
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
XX ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP007537.

XX 30-JUN-2000; 2000DE-01032529.

XX 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful for  
XX diagnosis and treatment of diseases associated with abnormal cytosine  
XX methylation.

XX Claim 1; SEQ ID NO 451; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
XX genes which are modified by the methylation of cytosines. The sequences  
XX can be used in the diagnosis and treatment of immune system disorders,  
XX including eye diseases such as retinopathy, neovascular glaucoma and  
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
XX diseases. The present sequence is a gene of the invention

XX Sequence 6593 BP; 1732 A; 121 C; 1448 G; 3292 T; 0 U; 0 Other;

Query Match 9.7%; Score 48; DB 6; Length 6593;

Best Local Similarity 45.8%; Pred. No. 0.018;

Matches 207; Conservative 0; Mismatches 240; Indels 5; Gaps 1;

QY 23 TTTGTTTGAATAGATCAGGGGTTTAGGCTTTTCCATTACTTTTAAAGTTTTCGT 82  
DB 274 TTTTGTGTTTATTTATGATGATTTTATGTTTATGTTTATGTTTATGTTTATGTTT 333  
QY 83 TACTGTCGCCGATCTGATTTTACGACATAGAGTTTCGGTTTTCGCCATCCAGTT 142  
DB 334 TTTTGTGTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 393  
QY 143 TGAATAAAGCTCCGCTTTTAAAGTTTGTGATCGATAAACCTGT-----GAAGATTG 197  
DB 394 TTAGATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 453  
QY 198 AGTCTAGTCGATTTATGATGATGATCCATCTTCATCGTTTTCCTGCTCGAAGTTCT 257

Db 454 TAAATGATTTGTTGGAGTTGATTAATTTTGTGTTGATTAAATTTGTTGTTAAATTT 513

Qy 258 GTATACACAGATTTGCTGTGTCGGAATGTCATACCTAGCCGTGATCGAGACTAGGG 317

Db 514 TTATGATGAATTTTATTTATTTAGTATTGTAATTTTATTTTATTTAGAAATTTTATTTGGTT 573

Qy 318 TTATTCGAGTCAATTTTGGCCCTTTTGGTTATATCTGTTTCGATAACGATTCATCTGGATT 377

Db 574 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 633

Qy 378 AGGGTTTAAAGTGGTACGCTTTTATGATTTCCAAATTTCTTCAAAATTTAGTTATGGAATG 437

Db 634 ATTTTGTGTTAGTTGTTTATTTGGGTTTTTTTGTAAATTTGTTGATTTATTTTAAAGATGAT 693

Qy 438 AATATCCCGAATGACGTTCATTTCTGTT 469

Db 694 ATTTTGAATTTTGTGTAGATTAATTTATAGAT 725

RESULT 10

ABL32638

ID AAS46530 standard; DNA; 7814 BP.

AC AAS46530;

DT 18-DEC-2001 (first entry)

DE Tumour suppressor gene derived chemically modified sequence #252.

XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;

KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;

KW cytosine methylation; ds.

XX Homo sapiens.

XX WO200168912-A2.

PD 20-SEP-2001.

PF 15-MAR-2001; 2001WO-EP002955.

PR 15-MAR-2000; 2000DE-01013847.

PR 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

PI WPI; 2001-602752/68.

DR

XX Fragments of chemically modified genes associated with tumor suppressor

PT genes and oncogenes, useful in designing primers and probes for analyzing

PT diseases associated with cytosine methylation state e.g. cancer.

XX

PS Claim 1; SEQ ID NO 252; 27pp; English.

XX

CC The invention relates to a nucleic acid comprising a sequence of 18

CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with

CC bisulphite, of genes associated with tumour suppression and oncogenes

CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and

CC 500 are missing from the sequence listing) sequences (Ss) and sequences

CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-

CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of

CC probes for detecting the cytosine methylation state and/or single

CC nucleotide polymorphisms and also to be used in an array for analysing

CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The

CC probes can also be used in a method for ascertaining genetic and/or

CC epigenetic parameters for the diagnosis and/or therapy of existing

CC diseases or the predisposition to specific diseases, by analysing

CC cytosine methylations. The parameters may be compared to another set of

CC genetic and/or epigenetic parameters, the differences serving as basis

CC for diagnosis and/or prognosis events which are disadvantageous to

CC patients. The present sequence is one of the 533 genomic sequences

CC derived from tumour suppressor genes and oncogenes. Sequences with even

CC numbered seq ID numbers are the complementary sequence of the

CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID

CC 535, except for those whose partner sequence is missing). Note: The

CC sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences

XX

Qy Sequence 7814 BP; 1677 A; 101 C; 1779 G; 4257 T; 0 U; 0 Other;

Query Match 9.4%; Score 46.6; DB 4; Length 7814;

Best Local Similarity 46.8%; Pred. No. 0.043;

Matches 214; Conservative 0; Mismatches 239; Indels 4; Gaps 2;

Qy 21 GATTTGTTTGTAAATTTAGATCAGGGGTTTAGGCTTTTCATTACTTTTAAATGTTTTTCT 80

Db 3356 GGTTTTTTATTTAGAGAGGGTTTGGTTTTATATTTTTTTTATTTAGTTTTTG 3415

Qy 81 GTTACTGTCCTCGGATCTGA-TTTTACGACAATAGAGTTTCGGGTTTTCCTCCATCCCA 139

Db 3416 ATTATTATATTAAATTTTATAGGGTTTTGTTTATTTAGTTTATTTTATTT 3475

Qy 140 GTTGAABAATAACGTCCTCTTTTAACTTTGCTGATCGATAAACCCTGCGAAGATTGAG 199

Db 3476 TTTTGGTAATGTTGAAATTAATTAATTTTGGTTTATTAATTTTTTTTGTATT 3535

Qy 200 TCTAGTCGATTTATTTGATGATCCATCTTCATCGTTTTTTCTTCGTCGAAATTCGT 259

Db 3536 TATTTATTTATTTGTTTTTTATGATTTATTTATTTGTTTATTTGTTTATTTAGTTATT 3595

Qy 260 ATAACAGATTTGCTGTGTCGATTTGTCATTCACCTGCGTATCGAGAACTAGGTT 319

Db 3596 ATTATTAAGTTTATTTATGATTAATTAATTTATGTTTATTTATTAATTTGTTGTT 3655

Qy 320 TTCGAGTCAATTTGCCCCCTTTTGGTTA---TATCTGGTTCGATAACGATTCATCTGGAT 376

Db 3656 GTTTATTTATTTGTTAGTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTT 3715

Qy 377 TAGGGTTTTTAAGTGGTGACGTTTAGTATCCAAATTTCTTCAAAATTTAGTTATGATAAT 436

Db 3716 AATTTATTTATGTTTATTTAAATTTAGTTCGTTTATTTGTTTATTTAGTTTATTAAT 3775

Qy 437 GAAATCCCGAATTTGACTGTTCAATTTCTTGTAAAT 473

Db 3776 TTTATTTGTTTATTTATTTATGAAATTTGTTTATTTAT 3812

## RESULT 11

ABL32638

ID ABL32638 standard; DNA; 6650 BP.

AC ABL32638;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 611.

XX Human; immune system disease; cytosine methylation; antiasthmatic;

XX antiarteriosclerotic; antianaemic; cytostatic; neotropic;

XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;

XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;

XX ds.

XX Homo sapiens.

OS WO200200928-A2.

XX

XX

XX

XX

XX

XX

XX

PD 03-JAN-2002.  
XX  
XX  
XX 02-JUL-2001; 2001WO-EP007537.  
XX  
XX 30-JUN-2000; 2000DE-01032529.  
PR 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.  
XX  
XX Claim 1; SEQ ID NO 611; 32pp + Sequence Listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention  
XX  
XX Sequence 6650 BP; 1628 A; 159 C; 1561 G; 3302 T; 0 U; 0 Other;  
SQ  
Query Match 9.4%; Score 46.4; DB 6; Length 6650;  
Best Local Similarity 44.0%; Pred. No. 0.047;  
Matches 197; Conservative 0; Mismatches 251; Indels 0; Gaps 0;  
QY 23 TTTGTTTGAATAGATCAGGGGTTAGGTCTTCCACTACTTTTAAATGTTTCTGT 82  
Db 3598 TGTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3657  
QY 83 TACTGTCGCGATCGATTTTACGACATAGAGTTTCGGGTTTGTCCCATTCAGTT 142  
Db 3658 TTATTAGTTTGTAGTTTGTAGTTTATATATAAGTGTTCGGTATTGTTGATGA 3717  
QY 143 TGAATATAACGTCGTTTAAAGTTTCTGTCGATCGATAAACCTCTGAAGATTGAGTCT 202  
Db 3718 TGATATTAGATGTTTGTGTAAGTTTAAATTTGTTAATTTATTTATTTAAT 3777  
QY 203 AGTCGATTTATGGATGATCCATCTTCAATCGTTTCTTCTGCTCGAAGTTCTGTATA 262  
Db 3778 ACCGAATGAGAGTCGAGTATGATGATGATGATGATGATGATGATGATGATGAT 3837  
QY 263 ACCAGATTGCTGTCGCGATTTGATACCTACCTAGCCGTCATCGAAGACTAGGGTTTC 322  
Db 3838 GTTTTATATAATTTAGATGATGATGATGATGATGATGATGATGATGATGATGAT 3897  
QY 323 GAGTCATTTTGGCCCTTTTGGTTATATCTGTTTTCGATACGATTCATCTGGAATAGGGT 382  
Db 3898 TATTTTATTTAGGGTTTATGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGAAT 3957  
QY 383 TTTAAGTGGGACGTTTGTAGTATCCCAATTTCTCAAAATTTAGTTATGGAATGAAT 442  
Db 3958 TCGTTTGTGATGTTGTGATGTTGTGATGTTGTGATGTTGTGATGTTGTGATGTTGGAAT 4017  
QY 443 CCCGAATTGACTGTTCAATTTCTTGTTA 470  
Db 4018 ATTTGGTTGATGTTGGAATTTTATTTT 4045  
RESULT 12  
ID ABL32509  
XX ABL32509 standard; DNA; 6048 BP.  
AC ABL32509;  
XX

DT 26-MAR-2002 (first entry)  
XX Human immune system associated gene SEQ ID NO: 482.  
DE  
XX  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; anti-anaemic; cytosinatic; nootropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
KW ds.  
XX  
XX Homo sapiens.  
OS  
XX W0200200928-A2.  
PN  
XX  
XX 03-JAN-2002.  
PD  
XX  
XX 02-JUL-2001; 2001WO-EP007537.  
PF  
XX 30-JUN-2000; 2000DE-01032529.  
PR 01-SEP-2000; 2000DE-01043826.  
XX  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-130909/17.  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful for  
PT diagnosis and treatment of diseases associated with abnormal cytosine  
PT methylation.  
XX  
XX Claim 1; SEQ ID NO 482; 32pp + Sequence Listing; German.  
XX  
XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention  
XX  
XX Sequence 6048 BP; 1296 A; 103 C; 1361 G; 3288 T; 0 U; 0 Other;  
SQ  
Query Match 9.4%; Score 46.2; DB 6; Length 6048;  
Best Local Similarity 46.9%; Pred. No. 0.051;  
Matches 144; Conservative 0; Mismatches 163; Indels 0; Gaps 0;  
QY 163 TTAAGTTTGTGATCGATAAACCTCTGACAGATTGAGTCTAGTCTATTTGATGATC 222  
Db 2325 TTAAGTTTGTGATCGATGATTAAGGTTTATAAATTTGATTTTATTTTGT 2384  
QY 223 CATCTTCATCGTTTCTTCTGCTCGAAGTTCTGTATACACAGATTGCTGTGCG 282  
Db 2385 GATTTTCTTGTATTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2444  
QY 283 ATTGTCATTACTAGCCGCTGATCGAAGCTAGGTTTTCGAGTCAATTTTCCCTTTT 342  
Db 2445 TTTTGAATTTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 2504  
QY 343 GGTATATCTGTTTCGATACGATTCATCTCGATTTAGGTTTAAAGTGTGACGTTAGT 402  
Db 2505 TTGGATTTTCTTCTGTAATAGTATGATGATGATGATGATGATGATGATGATGAT 2564  
QY 403 ATTCCAAATTTCTCAAAATTTAGTTATGATGATAAATCCGAATTCGACTGTCAAT 462  
Db 2565 GTTTAAATTTTATTTTATTTGAGGTTTCTTCTGAGTATTTTAAATTTGTAAGT 2624  
QY 463 TCTTGT 469  
|||

Detecting and differentiating between colon cell proliferative disorders associated with a gene or its regulatory regions comprises contacting a target nucleic acid in a biological sample obtained from the subject with a reagent.

Claim 32; SEQ ID NO 275; 74pp; English.

The invention relates to a novel method for detecting and differentiating between colon cell proliferative disorders associated with at least one gene or its regulatory regions. The method comprises contacting a target nucleic acid in a biological sample obtained from the subject with at least one reagent or a series of reagents, where the reagent or series of reagents, distinguishes between methylated and non methylated CpG dinucleotides within the target nucleic acid. The molecules of the invention demonstrate cytosinatic activity whilst the method may be useful for detecting and differentiating between colon cell proliferative disorders, including cancers such as colon adenoma and colon carcinoma. The PNA (peptide nucleic acid)-oligonomers are useful as probes for determining cytosine methylation state or single nucleotide polymorphisms. The current sequence is that of the pretreated genomic DNA

2625 TTTTTF 2631

RESULT 13  
Z10189  
AB210189 standard; DNA; 3025 BP.  
AB210189;  
16-JAN-2003 (first entry)  
Haematopoietic cell proliferation disorder related DNA sequence #329.  
Human; haematopoietic cell proliferation disorder; cytostatic;  
gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
cytosine methylation state; gene; ds.  
Homo sapiens.  
WO20027272-A2.  
03-OCT-2002.  
26-MAR-2002; 2002WO-EP003401.  
26-MAR-2001; 2001US-0278333P.  
(EPIG-) EPIGENOMICS AG.  
Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;  
Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;  
Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;  
Schwope I, Ziebarth H;  
WPI; 2003-018942/01.  
Detecting and differentiating between hematopoietic cell proliferative  
disorders, comprises contacting a target nucleic acid with a reagent that  
distinguishes between methylated and non-methylated CpG dinucleotides.  
Claim 28; SEQ ID NO 329; 117pp; English.  
The present invention describes a method for detecting and  
differentiating between haematopoietic cell proliferative disorders  
associated with at least 1 gene and/or their regulatory regions in a  
subject. The method comprises contacting a target nucleic acid in a  
biological sample obtained from the subject with at least 1 reagent,  
which distinguishes between methylated and non-methylated CpG  
dinucleotides within the target nucleic acid. AB209861 to AB21118  
represent specifically claimed nucleotide sequences from the present  
invention. Oligonucleotides from the present invention can be used; for  
differentiating between healthy haematopoietic cells and proliferative  
disorder haematopoietic cells; for differentiating between acute  
lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
determining the cytosine methylation state and/or single nucleotide  
polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
related sequences and their complements; and as primers for the  
amplification of haematopoietic cell proliferation disorder related DNA  
sequences. The nucleotide sequences from the present invention can also  
be used for detecting a predisposition to, differentiation between  
subclasses, diagnosis, prognosis, treatment and/or monitoring of  
haematopoietic cell proliferative disorders. The present method enables a  
highly specific classification of haematopoietic cell proliferative  
disorders allowing for improved and informed treatment of patients

Sequence 3025 BP; 535 A; 0 C; 930 G; 1560 T; 0 U; 0 Other;  
Query Match 9.3%; Score 46; DB 7; Length 3025;  
Best Local Similarity 47.9%; Pred. No. 0.049;  
Matches 162; Conservative 0; Mismatches 175; Indels 1; Gaps 1

23 TTTCTTTGTAATGATCAGCGGGTTAGGCTTTTCCATTACTTTTAAATGTTTTTCTGT 82

CC region of the invention. This sequence is not shown within the  
CC specification but is taken from Wipoweb.

XX SQ Sequence 3025 BP; 535 A; 0 C; 930 G; 1560 T; 0 U; 0 Other;

Query Match 9.3%; Score 46; DB 9; Length 3025;  
Best Local Similarity 47.9%; Pred. No. 0.049;  
Matches 162; Conservative 0; Mismatches 175; Indels 1; Gaps 1;  
23 TTGTTGTTGTAATTAGATCAGGGGTTTAGGTCCTTCCATTACTTTTAAATGTTTCTGT 82  
1881 TTGTTTGGTTTATTTAGGATGTTTATTTATTTATATATATGTTTATGTTGGGA 1940  
QY 83 TACTGTCGCCGATCTGATTTAGGCAATAGAGTTTCGGTTCCTCCATCCAGTT 142  
DB 1941 AGTGGTGTAGGAGATGAATATTAAGTAAAG-TATTGGGTTGGGGTGGGAGG 2059  
QY 143 TGAATAAAGCTCGCTCTTTTAAAGTTTTCGATGATGATGATGATGATGATGAT 202  
DB 2001 TTTAATTAAGTTTATGTTTAAATTTTAAAG-TATTGGGTTGGGGTGGGAGG 2059  
QY 203 AGTCGATTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 262  
DB 2060 TTTGTTGTTGTTTATTTTGTATTAATTAAGTTTTCCTCCATCCAGTT 2119  
QY 263 ACCGATTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 322  
DB 2120 GGAAGAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2179  
QY 323 GAGTCAATTTTCCCTTTTGGTTTATATCTGATGATGATGATGATGATGATGATGAT 360  
DB 2180 TTGAGAGTTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2217

RESULT 15  
ADEB4157  
ID ADEB4157 standard; DNA; 3025 BP.  
XX ADEB4157;  
XX 29-JAN-2004 (first entry)  
XX Human lymphoid cell proliferative disorder gene derived DNA #93.

XX ds; lymphoid cell proliferative disorder; methylation;  
KW methylated CpG dinucleotide; single nucleotide polymorphism; SNP;  
KW diffuse large B-cell lymphoma; mantle cell lymphoma;  
KW chronic lymphocytic leukemia; small lymphocytic lymphoma;  
KW follicular lymphoma; diagnosis; prognosis.

XX Homo sapiens.  
XX WO2003044226-A2.  
XX 30-MAY-2003.  
XX 25-NOV-2002; 2002WO-EP013265.  
XX 23-NOV-2001; 2001DE-01057491.  
XX 28-DEC-2001; 2001DE-01064501.  
XX (EPIG-) EPIGENOMICS AG.

XX Burger M, Caldwell C, Genc B, Becker E, Maier S, Nimmrich I;  
XX WPI; 2003-457621/43.

XX Detecting and differentiating between lymphoid cell proliferative  
PT disorders comprises contacting a target nucleic acid with at least one  
PT reagent that distinguishes between methylated and non-methylated CpG  
XX dinucleotides.

XX Claim 26; SEQ ID NO 153; 448pp; English.

XX  
CC The invention relates to a method of detecting and differentiating  
CC between lymphoid cell proliferative disorders associated with at least  
CC one gene and/or their regulatory regions in a subject by contacting a  
CC target nucleic acid in a biological sample obtained from the subject with  
CC at least one reagent or series of reagents that distinguish between  
CC methylated and non-methylated CpG dinucleotides within the target nucleic  
CC acid. The genes and/or their regulatory regions are preferably selected  
CC from MDR1, CSNK2B, EGR4, AR, CDK4, RB2, CDC25A, GP1b beta, MYOD1, CDH3,  
CC MYCL1, ELK1, ABL1, APC, BCL2, CDH1, CDKN1A, CDKN1B, CDKN2a, CDKN2B, FOS,  
CC GSTP1, HIC-1, MGMT, MHL1, MSL, MYC, MYC, MYC, MYC, MYC, MYC, MYC,  
CC GSK3beta, ESRL1, AFAF1, BAK1, BAX or HOXA5. Oligomers, peptide nucleic  
CC acid (PNA) oligomers and/or isolated nucleic acids based on the sequences  
CC of the genes are useful for detecting the methylation state of all the  
CC CpG dinucleotides within one or more the sequences, or their complements,  
CC for determining the cytosine methylation state and/or single nucleotide  
CC polymorphisms (SNPs) and for differentiating at least two of the medical  
CC conditions such as diffuse large B-cell lymphoma, mantle cell lymphoma,  
CC chronic lymphocytic leukemia, small lymphocytic lymphoma and follicular  
CC lymphoma. They are also useful for detecting of a predisposition to,  
CC differentiation between subclasses, diagnosis, prognosis, treating and/or  
CC monitoring of lymphoid cell proliferative disorder. This sequence  
CC represents a nucleic acid of a pretreated genomic DNA derived from the  
CC above mentioned genes.

XX SQ Sequence 3025 BP; 535 A; 0 C; 930 G; 1560 T; 0 U; 0 Other;

Query Match 9.3%; Score 46; DB 9; Length 3025;  
Best Local Similarity 47.9%; Pred. No. 0.049;  
Matches 162; Conservative 0; Mismatches 175; Indels 1; Gaps 1;

QY 23 TTGTTTGTAAATTAGATCAGGGGTTTAGGTCCTTCCATTACTTTTAAATGTTTCTGT 82  
DB 1881 TTGTTTGGTTTATTTAGGATGTTTATTTATTTATATATATGTTTATGTTTGGGA 1940  
QY 83 TACTGTCGCCGATCTGATTTTACGACATAGAGTTTCGGTTCCTCCATCCAGTT 142  
DB 1941 AGTGGTGTAGGAGATGAATATTAAGTAAAGTACTTTTCTGTTTGTGTTGGAT 2000  
QY 143 TGAATAAAGCTCGCTCTTTTAAAGTTTTCGATGATGATGATGATGATGATGATGAT 202  
DB 2001 TTTAATTAAGTTTATGTTTAAATTTTAAAG-TATTGGGTTGGGGTGGGAGG 2059  
QY 203 AGTCGATTTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 262  
DB 2060 TTGTTGTTGTTTATTTTGTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2119  
QY 263 ACCAGATTTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 322  
DB 2120 GGAAGAGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2179  
QY 323 GAGTCAATTTTCCCTTTTGGTTTATATCTGATGATGATGATGATGATGATGATGAT 360  
DB 2180 TTGAGAGTTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2217

Search completed: June 20, 2004, 01:58:58  
Job time : 368.167 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 00:55:23 ; Search time 3463.42 Seconds  
(without alignments)  
6182.178 Million cell updates/sec

Title: US-09-000-062-7  
Perfect score: 494  
Sequence: 1 ctcaggcgaagaacaggat.....cgcagatccggatctgcg 494

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sv.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_on.\*
- 21: em\_ov.\*
- 22: em\_pat.\*
- 23: em\_ph.\*
- 24: em\_pl.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rtd.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID        | Description        |
|------------|-------|---------------|--------|----|-----------|--------------------|
| 1          | 494   | 100.0         | 494    | 6  | AR182672  | AR182672 Sequence  |
| 2          | 490.8 | 99.4          | 494    | 6  | A59350    | A59350 Sequence 7  |
| 3          | 477.8 | 96.7          | 4833   | 8  | ATH3G     | X60429 A.thaliana  |
| C 4        | 477.8 | 96.7          | 112067 | 8  | ATCHRIV92 | AL161596 Arabidops |
| C 5        | 477.8 | 96.7          | 118267 | 8  | ATJSU17   | AD035708 Arabidops |
| 6          | 476.2 | 96.4          | 777    | 8  | AJ592632  | AJ592632 Arabidops |
| 7          | 55.8  | 11.3          | 12870  | 6  | AX348370  | AX348370 Sequence  |
| 8          | 55.8  | 11.3          | 12870  | 6  | AX348661  | AX348661 Sequence  |
| C 9        | 50.6  | 10.2          | 14281  | 8  | SKORFS    | X54850 S.klyverii  |
| 10         | 49.6  | 10.0          | 7218   | 6  | I66494    | I66494 Sequence 14 |
| 11         | 49.6  | 10.0          | 34980  | 6  | AX344573  | AX344573 Sequence  |
| 12         | 49    | 9.9           | 17869  | 6  | AX345007  | AX345007 Sequence  |
| 13         | 49    | 9.9           | 17869  | 6  | AX348307  | AX348307 Sequence  |
| C 14       | 49    | 9.9           | 176162 | 5  | AL953898  | AL953898 Zebrafish |
| 15         | 48.6  | 9.8           | 250029 | 3  | A2014820  | AE014820 Plasmodi  |
| 16         | 48    | 9.7           | 6593   | 6  | AX345380  | AX345380 Sequence  |
| 17         | 47.8  | 9.7           | 93791  | 2  | AC138073  | AC138073 Homo sapi |
| C 18       | 47.2  | 9.6           | 165888 | 2  | AC144291  | AC144291 Macaca mu |
| C 19       | 47.2  | 9.6           | 166224 | 2  | AC142946  | AC142946 Macaca mu |
| 20         | 47    | 9.5           | 1453   | 8  | AJ591978  | AJ591978 Arabidops |
| 21         | 47    | 9.5           | 95764  | 2  | AC015732  | AC015732 Homo sapi |
| 22         | 46.6  | 9.4           | 7814   | 6  | AX251284  | AX251284 Sequence  |
| 23         | 46.4  | 9.4           | 6650   | 6  | AX345540  | AX345540 Mus muscu |
| 24         | 46.4  | 9.4           | 37159  | 2  | AC100259  | AC100259 Mus muscu |
| 25         | 46.4  | 9.4           | 56833  | 2  | AC084077  | AC084077 Homo sapi |
| 26         | 46.2  | 9.4           | 6048   | 6  | AX345411  | AX345411 Sequence  |
| C 27       | 46.2  | 9.4           | 164103 | 2  | AC037484  | AC037484 Homo sapi |
| C 28       | 46.2  | 9.4           | 190701 | 9  | AC022274  | AC022274 Homo sapi |
| 29         | 46    | 9.3           | 3025   | 6  | AX598989  | AX598989 Sequence  |
| 30         | 46    | 9.3           | 3025   | 6  | AX767505  | AX767505 Sequence  |
| 31         | 46    | 9.3           | 3025   | 6  | AX795863  | AX795863 Sequence  |
| 32         | 46    | 9.3           | 3025   | 6  | AX822383  | AX822383 Sequence  |
| 33         | 46    | 9.3           | 3025   | 6  | AX826023  | AX826023 Sequence  |
| C 34       | 46    | 9.3           | 151900 | 2  | BX537336  | BX537336 Danio rer |
| 35         | 45.8  | 9.3           | 165398 | 2  | AC104677  | AC104677 Sus scro  |
| 36         | 45.6  | 9.2           | 6668   | 6  | AX346599  | AX346599 Sequence  |
| 37         | 45.6  | 9.2           | 7442   | 6  | AX251441  | AX251441 Sequence  |
| 38         | 45.6  | 9.2           | 66993  | 2  | AC138074  | AC138074 Homo sapi |
| 39         | 45.6  | 9.2           | 170827 | 2  | AC125567  | AC125567 Rattus no |
| 40         | 45.6  | 9.2           | 236566 | 2  | AC094427  | AC094427 Rattus no |
| 41         | 45.4  | 9.2           | 164535 | 2  | CNS05TCJ  | AL355100 Homo sapi |
| 42         | 45.2  | 9.1           | 88549  | 3  | AC116979  | AC116979 Dictyoste |
| C 43       | 45    | 9.1           | 127709 | 2  | CNS05TC1  | AL352983 Homo sapi |
| C 44       | 45    | 9.1           | 178948 | 5  | AL929434  | AL929434 Zebrafish |
| 45         | 44.8  | 9.1           | 5208   | 6  | AX345825  | AX345825 Sequence  |

ALIGNMENTS

|            |  |                                    |        |     |        |                 |
|------------|--|------------------------------------|--------|-----|--------|-----------------|
| RESULT 1   | AR182672   | AR182672                           | 494 bp | DNA | linear | PAT 20-APR-2002 |
| LOCUS      | AR182672   | Sequence 7 from patent US 6338961. |        |     |        |                 |
| DEFINITION | AR182672   |                                    |        |     |        |                 |
| ACCESSION  | AR182672   |                                    |        |     |        |                 |
| VERSION    | AR182672.1   | GI:20225879                        |        |     |        |                 |
| KEYWORDS   |  |                                    |        |     |        |                 |
| SOURCE     | Unknown.   |                                    |        |     |        |                 |
| ORGANISM   | Unknown.   |                                    |        |     |        |                 |
| REFERENCE  | 1 (bases 1 to 494)   |                                    |        |     |        |                 |
| AUTHORS    | Derose, R., Chaubet, N. and Gigot, C.  |                                    |        |     |        |                 |
| TITLE      | Isolated DNA sequence capable of serving as regulatory element in a chimeric gene which can be used for the transformation of plants |                                    |        |     |        |                 |
| JOURNAL    | Patent: US 6338961-A 7 15-JAN-2002;  |                                    |        |     |        |                 |

```

1.103
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"

```

```
/db xref="taxon:3702"
/cell_line="Columbia"
/clone_lib="cosmid c22"
/dev_stage="callus"
691..1100
/gene="H3, gene1"
<691
/gene="H3, gene1"
/number=1
692..1100
/gene="H3, gene1"
/number=1
join(1101..1172,1260..1336,1431..1521,1620..1790)
/gene="H3, gene1"
join(1101..1172,1260..1336,1431..1521,1620..1790)
/gene="H3, gene1"
/codon_start=1
/product="histone H3.3 like protein"
/protein_id="CAA42957.1"
/db xref="GI:16324"
/db xref="GOA:P59169"
/db xref="SWISS-PROT:P59169"
/translation="MARTKQTARKSTGGKAPRQLATKAARKSAPTTGGVKKPHRYRP
GTVALRIRKYQKSTELLIRKLPQRLVREIAQDFKTDLRFQSHAVIALQEAAREALV
GLFEDNLCAIHAKRVITMPKDIQLARRIGERA"
1101..1172
/gene="H3, gene1"
1101..1172
/gene="H3, gene1"
/number=2
1173..1259
/gene="H3, gene1"
1173..1259
/gene="H3, gene1"
/number=2
1260..1336
/gene="H3, gene1"
1260..1336
/gene="H3, gene1"
/number=3
1337..1430
/gene="H3, gene1"
1337..1430
/gene="H3, gene1"
/number=3
1431..1521
/gene="H3, gene1"
1431..1521
/gene="H3, gene1"
/number=4
1522..1619
/gene="H3, gene1"
1522..1619
/gene="H3, gene1"
/number=4
1620..1790
/gene="H3, gene1"
1620..1790
/gene="H3, gene1"
/number=5
join(2912..2983,3065..3141,3227..3488)
/gene="H3, gene 2"
join(2912..2983,3065..3141,3227..3488)
/gene="H3, gene 2"
/codon_start=1
/product="histone H3.3 like protein"
/protein_id="CAA42958.1"
/db xref="GI:404825"
/db xref="GOA:P59169"
/db xref="SWISS-PROT:P59169"
/translation="MARTKQTARKSTGGKAPRQLATKAARKSAPTTGGVKKPHRYRP
GTVALRIRKYQKSTELLIRKLPQRLVREIAQDFKTDLRFQSHAVIALQEAAREALV
GLFEDNLCAIHAKRVITMPKDIQLARRIGERA"
```

```
gene 2912..2983
/exon /gene="H3, gene 2"
<2912..2983
/exon /gene="H3, gene 2"
/number=1
2984..3064
/exon /gene="H3, gene 2"
2984..3064
/exon /gene="H3, gene 2"
/number=1
3065..3141
/exon /gene="H3, gene 2"
3065..3141
/exon /gene="H3, gene 2"
/number=2
3142..3226
/exon /gene="H3, gene 2"
3142..3226
/exon /gene="H3, gene 2"
/number=2
3227..3488
/exon /gene="H3, gene 2"
3227..3488
/exon /gene="H3, gene 2"
/number=3

ORIGIN
Query Match 96.7%; Score 477.8; DB 8; Length 4833;
Best Local Similarity 99.6%; Pred. No. 1.6e-101;
Matches 479; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGCGAAGAACAGGATGATTTGTTGTAATAGATCAGGGGTTTAGTCTTTCCAT 60
DB 2433 CTCAGCGAAGAACAGGATGATTTGTTGTAATAGATCAGGGGTTTAGTCTTTCCAT 2492
QY 61 TACTTTTTTAATGTTTTTCTGTTACTGTCCTCGGATCTGATTTTACGACAAATAGAGTTT 120
DB 2493 TACTTTTTTAATGTTTTTCTGTTACTGTCCTCGGATCTGATTTTACGACAAATAGAGTTT 2552
QY 121 CGGGTTTTGCCATTCAGTTTGAAAAATAAACGTCGCTCTTTTAAAGTTTCTCGATCGA 180
DB 2553 CGGGTTTTGCCATTCAGTTTGAAAAATAAACGTCGCTCTTTTAAAGTTTCTCGATCGA 2612
QY 181 TAAACCTGTGAAGTTGATCTAGTCGATTTTATTTGGATGATCCATTTCTTCATCGTTT 240
DB 2613 TAAACCTGTGAAGTTGATCTAGTCGATTTTATTTGGATGATCCATTTCTTCATCGTTT 2672
QY 241 TCTTGCCTTCGAAGTTCTGTATAACAGATTTTGTCTGTGTGGATTGTCAATACCTAGCCG 300
DB 2673 TCTTGCCTTCGAAGTTCTGTATAACAGATTTTGTCTGTGTGGATTGTCAATACCTAGCCG 2732
QY 301 TGTATCGAAGACTAGGTTTTCGAGTCAATTTTGGCCCTTTTGGTTATATCTGTTCCAT 360
DB 2733 TGTATCGAAGACTAGGTTTTCGAGTCAATTTTGGCCCTTTTGGTTATATCTGTTCCAT 2792
QY 361 AACGATTCACTCGGATTAGGGTTTAAAGTGTGAGTTAGTTTCAATTTCTTCAAAA 420
DB 2793 AACGATTCACTCGGATTAGGGTTTAAAGTGTGAGTTAGTTTCAATTTCTTCAAAA 2852
QY 421 TTTAGTTATGGATAATGAAATCCGAAATGACCTGTTTCAATTTCTTGTAAATGGCGAGA 480
DB 2853 TTTAGTTATGGATAATGAAATCCGAAATGACCTGTTTCAATTTCTTGTAAATGGCGAGA 2912
QY 481 T 481
DB 2913 T 2913

RESULT 4
ATCHRIV92/c 112067 bp DNA linear PLN 16-MAR-2000
LOCUS Arabidopsis thaliana DNA chromosome 4, contig fragment No. 92.
DEFINITION
ACCESSION AL161596
```

|           |  |   |        |
|-----------|--|---|--------|
| VERSION   | AL161596.2   | GI:7271037  |        |
| KEYWORDS  | Arabidopsis thaliana (thale cress)   |   |        |
| SOURCE    | Arabidopsis thaliana   |   |        |
| ORGANISM  | Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.   |   |        |
| REFERENCE | 1 (bases 1 to 107700)  | Murphy, G., Ridley, P., Hudson, S., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.   | intron |
| AUTHORS   | Unpublished  |   |        |
| JOURNAL   | 2 (bases 9546 to 9977)   | Volckaert, G., Grymonprez, B., Voet, M., Robben, J., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.  | exon   |
| REFERENCE | 3 (bases 107578 to 112067)   | Rose, M., Hempel, S., Entian, K.-D., Mewes, H.W., Lemcke, K. and Mayer, K.F.X.  | intron |
| AUTHORS   | Unpublished  |   |        |
| JOURNAL   | 4 (bases 1 to 112067)  | EU Arabidopsis sequencing, project.   | exon   |
| REFERENCE | Submitted (10-MAR-2000)  | MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de | gene   |
| AUTHORS   | Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk   |   | CDS    |
| JOURNAL   | Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> and this fragment has an overlap with ATCHRIV91 at the 5' end and an overlap with ATCHRIV93 at the 3' end.   |   |        |
| COMMENT   | Location/Qualifiers  |   |        |
| FEATURES  | 1..112067  |   |        |
| source    | /organism="Arabidopsis thaliana"   |   |        |
|           | /molecule="genomic DNA"  |   |        |
|           | /variety="Columbia"  |   |        |
|           | /db_xref="taxon:3702"  |   |        |
|           | /chromosome="4"  |   |        |
|           | 6122..8203   |   |        |
| gene      | /gene="AT4g39830"  |   |        |
| CDS       | join(6122..6415,6481..6573,6652..6864,6959..7153,7250..8203)   |   |        |
|           | /gene="AT4g39830"  |   |        |
|           | /note="strong similarity to L-ascorbate oxidase, Cucumis sativus, PIR:K5KVAO   |   |        |
|           | Contains Glycosyl hydrolases family 5 signature  |   |        |
|           | AA384-393; Multicopper oxidases signatures   |   |        |
|           | AA131-151; Multicopper oxidases signatures   |   |        |
|           | AA540-560; Multicopper oxidases signatures   |   |        |
|           | AA545-556; Prokaryotic membrane lipoprotein lipid attachment site AA567-577  |   |        |
|           | contains EST gb:AI994215.1, AA585817, AI996008.1   |   |        |
|           | /codon_start=1   |   |        |
|           | /product="putative L-ascorbate oxidase"  |   |        |
|           | /protein_id="CA380646.1"   |   |        |
|           | /db_xref="GI:7271038"  |   |        |
|           | /db_xref="GOA:O65670"  |   |        |
|           | /db_xref="SPTREMBL:O65670"   |   |        |
|           | /translation="MKRKRSDTVHVFNLMLVLCFTALFFSVLCQKIRRFKWEVKY EFKSPDFEKLITNGKPGFTIKQAQGDITVELKNSFMTENVAVHWHGIRQIGTP WFGVEGTCPLTPGEVFTYQFVDRPGTYMHSHYGMQRESGLIGMIQVFPATPE PTYDTRNFLTDYHVKMSKATGLASIPPKWYGEQPSLMIOGRGFNCSNLLTT PRLSEGVGNVADCSRFLLTVIPGKTVLRIGSLTALSAPQLEGNLFTVADG HYVEPPFVNKLVYSSETYSVLKADQNRPNRYWISSIVSRPPTPATAVINYFPN HPRRPPTSSNIVPEWNRSLAQSLAKARRGFIHALPNSDKIVILNLTONEV NGYRWNVSVYHHPKPLIALKQNLNADWRFAPENYDSRYNIDYIPAKPLNANA TTSDGYRLNFSTVILQNTMNNANSETHPMLHGHDFWVLGYGEGRNESEDP KRYNRVDPIKKNTVAVQPCFWGATLRADPNPGVWSPHCHIESHFFMGMGIVFESGIDK VSSLPSSIMCGGQTKR |   |        |
|           | 6122..6415   |   |        |
| exon      | /gene="AT4g39830"  |   |        |
|           | /number=1  |   |        |
|           | 6416..6480   |   |        |
|           | /gene="AT4g39830"  |   |        |
|           | /number=1  |   |        |
|           | 6481..6573   |   |        |
|           | /gene="AT4g39830"  |   |        |
|           | /number=2  |   |        |
|           | 6574..6651   |   |        |
|           | /gene="AT4g39830"  |   |        |
|           | /number=2  |   |        |
|           | 6652..6864   |   |        |
|           | /gene="AT4g39830"  |   |        |
|           | /number=3  |   |        |
|           | 6865..6958   |   |        |
|           | /gene="AT4g39830"  |   |        |
|           | /number=3  |   |        |
|           | 6959..7153   |   |        |
|           | /gene="AT4g39830"  |   |        |
|           | /number=4  |   |        |
|           | 7154..7249   |   |        |
|           | /gene="AT4g39830"  |   |        |
|           | /number=4  |   |        |
|           | 7250..8203   |   |        |
|           | /gene="AT4g39830"  |   |        |
|           | /number=5  |   |        |
|           | 12287..13642   |   |        |
|           | /gene="AT4g39840"  |   |        |
|           | 12287..13642   |   |        |
|           | /gene="AT4g39840"  |   |        |
|           | /note="contains EST gb:AI999725.1, Z46537, F13826, AI996886.1"   |   |        |
|           | /codon_start=1   |   |        |
|           | /product="putative protein"  |   |        |
|           | /protein_id="CA380647.1"   |   |        |
|           | /db_xref="GI:7271039"  |   |        |
|           | /db_xref="SPTREMBL:O9SMR9"   |   |        |
|           | /translation="MSPLKYLCKRVLLLVSLFLFSSSSFPASSVDDHELQRIIDP HFVRRLVLDLSTGDDDETNI PPPKKKLLITDSVSSSSSGTKNQTKLLKPIS SSSSTKQTKLAKTTTGSHKLNSTKSSNTTKTSSELKLNSTGKTKNQTKLLKPIS ADLSKSSSKNKTITPPSKLSPSPKSKSPKSPKSKSKSEKEIKPFWDDEED EDVSEFEDLPQFSLIPDLERISTTKNYINKANKQITKNKPFYFGNKVAPTAS VWFVFLVPLLVLIIFNRKAYFSIQKILIIQIYLSIYFSLICLSLVTGIEPLK FLVATSSSTVYLIQITGVYFVLLMLVLFSTDCGLGKVLGLAQTGVPVAV GLHYVAVFHRVLRQPPKTNWIKHGVYATCFLLICLSNAERRKXEYLEEGDEGKK N"   |   |        |
|           | 12287..13642   |   |        |
|           | /gene="AT4g39840"  |   |        |
|           | /number=1  |   |        |
|           | 15539..16205   |   |        |
|           | /note="EST N95877 matches to positions 26106-26355, 26723-26772; possibly related to subsequent gen"   |   |        |
|           | 16239..23312   |   |        |
|           | /gene="AT4g39850"  |   |        |
|           | join(16239..16300,16394..16679,16981..17205,17341..17539,17675..18129,18222..18357,18433..18511,19039..19271,19326..19630,19712..19844,19901..20056,20213..20435,20515..20623,20845..20935,21205..21372,21453..21568,21626..21860,21924..22104,22162..22362,22534..22680,22750..22983,23134..23312)  |   |        |
|           | /gene="AT4g39850"  |   |        |
|           | /note="similarity to 70K peroxisomal membrane protein, Rattus norvegicus, PIR: A35723  |   |        |
|           | Contains ABC transporters family signature AA580-594; ABC transporters family signature AA1369-1383; ATP/GTP-binding site motif A (P-loop) AA1247-1254   |   |        |
|           | contains EST gb:N95877"  |   |        |
|           | /codon_start=1   |   |        |
|           | /product="putative protein"  |   |        |
|           | /protein_id="CA380648.1"   |   |        |
|           | /db_xref="GI:7271040"  |   |        |
|           | /db_xref="GOA:O9SMR8"  |   |        |
|           | /db_xref="SPTREMBL:O9SMR8"   |   |        |
|           | /translation="WPSLQLQLQLTERGRLVASRRKSILLAAAGIVAAGGTAVYLKSRV  |   |        |

ASRPDSRLCNGSDDELEKLTATDQNAKITTKKKGGGLKSLQVLTAILLQMG  
KMGARDLLAVVVFRTALSNRLAKVQGFLEAAFLRRLPLRLSLNMLCPMLS  
TLHSTKYITGALSURFKRLKILHSHYFENWYIKKSHVDGRITHPQRLASDVR  
FSELSDLIIDDLTAVDGLYAWRLCSYASPKYIIFWILAYLGAGTARFNSPSPGK  
LMSEQLLEGEYQLHSLRTHSESAFYCGETREESHQOQKFNIVSMHSHVLDHW  
WFGMIQDLSKLYKATVAVLIIIEFPFSGHLRPDDSTLAGAEMLSNIRYHTSVISLF  
CALGTLISRRRLNRLSGADRIHEPMAVRELSDGDKSPFORNRNRYLSSEANYVEF  
SDVKNVPTGNVLEDLTAVEQSNLLITGGLWPLVSGHIVKPGVGSDLNKEIFVVP  
QRPMAVGTLDQIYPLTSGQSELLTELQMVELLKNVLSILVSEYTRWDLVLELD  
RYQKEVNWGDDELSGEQRLGWARLYHKPFAILDCTSAVTTIDMEERPAKVRA  
MSCSITISHRPALVAFHDVVLSDGEGGSVHYKDDSDALLTDAEISVKSDDTRQ  
NDAMVQVFAAARAAVEFLFESATNSKASQYQTLIARSPVVDKSVLPPFPQPT  
SQRALPSRVAMNLPLPKWPHLIFCISTWLTYSACADANIDILIKITDHFIVTQI  
PTIPDKQAOALLAVALCVLISRTLSIDRIASLNGTIVKYLEQDAAFLVRLIGLSVQS  
GASIIAPSRHLTORLALGWRILTOHLRLNRYLNNAFYKVFHMSGNSIDADQRLTR  
DLEKLTADLSGLTGWKVPVDILWYRYVGLISNICPREGVAILTYMLLGLGFLRR  
VAPFDGLAEEQLEQKFRPSTSCLCRFHERLNTHAESIAFFGGAREKAVFLIA  
LATAAGFWVIFESKPEVNELEFYHLCDDGLFMSNVASLMVDKKFRALLDHMLLR  
KMLYGLIDDDFTVKQLEPNVWGLSLLYALEHKGDRALVSTQVKYLFHLYMLLIPGE  
LAHALRYLASVQSFWAFGDILELHKFLLELGGINRIIFELDEFILDASQSGVTSNQ  
TSRLSDOLLFSFSVDIITPAQKMAKSLCEIVSGKSLVTPNGSGKTSVFRVLRD  
IWPVTCRLPKPSLDIKELASGNGMFFVPORPVTCGLTGLRDQIIYPLSKEAEKRAK  
LYTSGESTAGSILDSHLKILLENVRLVILLERDVGGWDATTNWDIISLGEQRLG  
MVSLL

exon 16239..16300  
/gene="AT4g39850"  
/number=1

intron 16301..16393  
/gene="AT4g39850"  
/number=1

exon 16394..16679  
/gene="AT4g39850"  
/number=2

intron 16680..16980  
/gene="AT4g39850"  
/number=2

exon 16981..17205  
/gene="AT4g39850"  
/number=3

intron 17206..17340  
/gene="AT4g39850"  
/number=3

exon 17341..17538  
/gene="AT4g39850"  
/number=4

intron 17539..17674  
/gene="AT4g39850"  
/number=4

exon 17675..18129  
/gene="AT4g39850"  
/number=5

Query Match 96.7%; Score 477.8; DB 8; Length 112067;  
Best Local Similarity 99.6%; Pred. No. 1.2e-101;  
Matches 479; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTCAGCGAAGAACAGGATGATTTGTTGTAATAGATCAGGGGTTTAGTCTTTCCAT 60  
DB 83915 CTCAGCGAAGAACAGGATGATTTGTTGTAATAGATCAGGGGTTTAGTCTTTCCAT 83956

QY 61 TACTTTTTAAATGTTTTTTCTGTTACTGTCCTCGGATCTGATTTTACGACATAGAGTTT 120  
DB 83855 TACTTTTTAAATGTTTTTTCTGTTACTGTCCTCGGATCTGATTTTACGACATAGAGTTT 83796

QY 121 CGGGTTTGTCCCATTCAGTTTGAAATAAAGCTCGCTCTTTTAAAGTTTCTCGATCGA 180  
DB 83795 CGGGTTTGTCCCATTCAGTTTGAAATAAAGCTCGCTCTTTTAAAGTTTCTCGATCGA 83736

QY 181 TAAACCTGTGAAGATTGAGTGTAGTCGATTTTATTTGGATGATCCATTTCTTCATCGTTTTT 240  
DB 83735 TAAACCTGTGAAGATTGAGTGTAGTCGATTTTATTTGGATGATCCATTTCTTCATCGTTTTT 83676

QY 241 TCTTCTCTCGAAGTTTCTGTATAACACAGATTGTCTGTGTGGCATGTGTCATTACCTAGCG 300

DB 83675 TCTTCTCTCGAAGTTTCTGTATAACACAGATTGTCTGTGTGGCATGTGTCATTACCTAGCG 83616

QY 301 TGTATCGAAGACTAGGGTTTTCGAGTCAATTTGCCCCCTTTTGTTTATATCTGGTTCCAT 360

DB 83615 TGTATCGAAGACTAGGGTTTTCGAGTCAATTTGCCCCCTTTTGTTTATATCTGGTTCCAT 83556

QY 361 AACGATTTCATCTCGAATTAGGGTTTAACTGGTGAGTTTAGTATTCCTCAATTTCTTCAAAA 420

DB 83555 AACGATTTCATCTCGAATTAGGGTTTAACTGGTGAGTTTAGTATTCCTCAATTTCTTCAAAA 83496

QY 421 TTTAGTTATGGATAATGAAATCCGAATGACGTGTTCAATTTCTTGTAAATGCGCAGA 480

DB 83495 TTTAGTTATGGATAATGAAATCCGAATGACGTGTTCAATTTCTTGTAAATGCGCAGA 83436

QY 481 T 481

DB 83435 T 83435

RESULT 5  
LOCUS ATTSJ17/c  
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone T5J17 [SSA project].  
ACCESSION AL035708  
VERSION AL035708.2 GI:5918309  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1  
AUTHORS Bevan, M., Murphy, G., Ridley, P., Hudson, S., Bancroft, I., Mewes, H.W., Mayer, K.F.X., Lemcke, K. and Schueller, C.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 118267)  
AUTHORS EU Arabidopsis sequencing, project.  
TITLE Direct Submission  
JOURNAL Submitted (21-SEP-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UU Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk

COMMENT On Sep 22, 1999 this sequence version replaced gi:4490734. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES  
source 1..118267  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
misc\_feature 1..22789  
/note="position 1-22789 overlaps to BAC T19P19, EMBL accession AL022605; please refer to this entry for analysis and annotation"  
gene 22854..24209  
/gene="T5J17.10"  
CDS 22854..24209  
/gene="T5J17.10"  
/note="contains EST gb:FL3826, 246537"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CA33897.1"  
/db\_xref="GI:4490735"  
/translation="MSPLKYLCFKRLVLLLYSSLFSSSSFFASSVDDHQLRIDP HFVRRLVQDLSDSGDDDETNLPPPKKKLLTDSVSSSSSSSGTCKKQTKLKPIS







## artificial sequences.

REFERENCE 1  
 AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
 TITLE Method and nucleic acids for pharmacogenomic methylation analysis  
 JOURNAL Patent: WO 0202806-A 65 10-JAN-2002;  
 Epigenomics AG (DE)  
 FEATURES  
 source Location/Qualifiers  
 1..12870  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="chemically treated genomic DNA (Homo sapiens)"

## ORIGIN

Query Match 11.3%; Score 55.8; DB 6; Length 12870;  
 Best Local Similarity 46.5%; Pred. No. 0.0061;  
 Matches 180; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

QY 18 TATGATTGTTGTAATTAGATCAGGGGTTTAGGTCCTTCCATTACTTTTAAATGTTTT 77  
 Db 3147 TTATTTGGTTTAAATGTTTATATTTTAAAGTTTATTAATGATGATGTTTTT 3206  
 QY 78 TCTGTTACTGTCCTCCGATCTGATTTTACGACATAGAGTTTCGGGTTTGTCCCATTC 137  
 Db 3207 TATAGTTTTTAAATAGTATTTGTTTTTATATGTTTATTAATGATGATGTTTTT 3266  
 QY 138 CAGTTTGAATAAAGCTCCGCTTTTAAAGTTTGTGATCGATAAACCTGTGAAGATTG 197  
 Db 3267 TATTGTTGTTTATTTATATTTGTTTTTTTTTTTGGAAATGTAATTCGGTTTTTTTT 3326  
 QY 198 AGTCTAGTCGATTTATGATGATCCATTCCTTCATCGTTTTTTCTTGCCTCGAAGTTCT 257  
 Db 3327 TTTTAGTTTTTATTTGTTTTTATTTTAAATTTATTTTTTTTAGGAAGTTTTT 3386  
 QY 258 GTATACCAAGATTGCTGTCGCGATTGCTACCTAGCCGTGATCGAAGCTAGGG 317  
 Db 3387 TTTTGATTTATTTAGTTTATTTTATTTATTTTATTTTGGATTTTTTTAGGAATTAAG 3446  
 QY 318 TTTTCGAGTCAATTTTGCCTTTGTTGTTATATCTGTTTCGATAACGATTCATCTGGATT 377  
 Db 3447 TTAGTGAATACGATTGATATTTTATAATAATTTATTTGGGTTTGGTTTTTTTTT 3506  
 QY 378 AGGTTTAAAGTCGTGACGTTTAGTAT 404  
 Db 3507 TAGAAGTCGGTAGGTAATGATGTTT 3533

RESULT 8  
 AX348661 12870 bp DNA linear PAT 06-FEB-2002  
 LOCUS Sequence 119 from Patent WO0202807.  
 DEFINITION AX348661  
 ACCESSION AX348661  
 VERSION AX348661.1 GI:18614696  
 KEYWORDS  
 SOURCE synthetic construct  
 ORGANISM synthetic construct  
 artificial sequences.

REFERENCE 1  
 AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
 TITLE Diagnosis of diseases associated with cell signalling  
 JOURNAL Patent: WO 0202807-A 119 10-JAN-2002;  
 Epigenomics AG (DE)

FEATURES  
 source Location/Qualifiers  
 1..12870  
 /organism="synthetic construct"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:32630"  
 /note="chemically treated genomic DNA (Homo sapiens)"

## ORIGIN

Query Match 11.3%; Score 55.8; DB 6; Length 12870;  
 Best Local Similarity 46.5%; Pred. No. 0.0061;  
 Matches 180; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

QY 18 TATGATTGTTGTAATTAGATCAGGGGTTTAGGTCCTTCCATTACTTTTAAATGTTTT 77  
 Db 3147 TTATTTGGTTTAAATGTTTATATTTTAAAGTTTATTAATGATGATGTTTTT 3206  
 QY 78 TCTGTTACTGTCCTCCGATCTGATTTTACGACATAGAGTTTCGGGTTTGTCCCATTC 137  
 Db 3207 TATAGTTTTTAAATAGTATTTGTTTTTATATGTTTATTAATGATGATGTTTTT 3266  
 QY 138 CAGTTTGAATAAAGCTCCGCTTTTAAAGTTTGTGATCGATAAACCTGTGAAGATTG 197  
 Db 3267 TATTGTTGTTTATTTATATTTGTTTTTTTTTTGGAAATGTAATTCGGTTTTTTTT 3326  
 QY 198 AGTCTAGTCGATTTATGATGATCCATTCCTTCATCGTTTTTTCTTGCCTCGAAGTTCT 257  
 Db 3327 TTTTAGTTTTTATTTGTTTTTATTTTAAATTTATTTTTTTTAGGAAGTTTTT 3386  
 QY 258 GTATACCAAGATTGCTGTCGCGATTGCTACCTAGCCGTGATCGAAGCTAGGG 317  
 Db 3387 TTTTGATTTATTTAGTTTATTTTATTTATTTTGGATTTTTTTAGGAATTAAG 3446  
 QY 318 TTTTCGAGTCAATTTTGCCTTTGTTGTTATATCTGTTTCGATAACGATTCATCTGGATT 377  
 Db 3447 TTAGTGAATACGATTGATATTTTATAATAATTTATTTGGGTTTGGTTTTTTTTT 3506  
 QY 378 AGGTTTAAAGTCGTGACGTTTAGTAT 404  
 Db 3507 TAGAAGTCGGTAGGTAATGATGTTT 3533

## RESULT 9

SKORPS/c 14281 bp DNA linear PLN 06-NOV-1996  
 LOCUS S.kluveri linear plasmid pSKL DNA for open reading frames 1-10.  
 DEFINITION  
 ACCESSION X54850  
 VERSION X54850.1 GI:4868  
 KEYWORDS linear plasmid; plasmid.  
 SOURCE Saccharomyces kluyveri  
 ORGANISM Saccharomyces kluyveri  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Saccharomyces.

## REFERENCE

AUTHORS Hishinuma, F. and Hirai, K.  
 TITLE Genome organization of the linear plasmid, pSKL, isolated from  
 Saccharomyces kluyveri  
 JOURNAL Mol. Gen. Genet. 226 (1-2), 97-106 (1991)  
 MEDLINE 91238725  
 PUBMED 2034232  
 REFERENCE 2 (Bases 1 to 14281)  
 Hishinuma, F.  
 Direct Submission  
 Submitted (17-OCT-1990) Hishinuma F., Mitsubishi Kasei Institute of  
 Life Sciences, 11 Minamiooya Machida-shi, Tokyo 194, Japan

## FEATURES

source Location/Qualifiers  
 1..14281  
 /organism="Saccharomyces kluyveri"  
 /mol\_type="genomic DNA"  
 /strain="UVC40"  
 /db\_xref="taxon:4934"  
 /plasmid="pSKL"  
 1..483  
 /note="terminal"  
 /rpt\_type=INVERTED  
 558..51279  
 /note="ORF 1"  
 578..1279  
 /note="unnamed protein product; ORF 1"  
 /codon\_start=1  
 /protein\_id="CAA38620.1"  
 /db\_xref="GI:4868"  
 /db\_xref="GOA:Q04318"  
 /db\_xref="SPTREMBL:Q04318"  
 /translation="MEMDRLFSAYSGDYKTYISGENSESELEVAICIKNGCFINIK"

## repeat\_unit

repeat\_unit  
 /note="terminal"  
 /rpt\_type=INVERTED  
 558..51279  
 /note="ORF 1"  
 578..1279  
 /note="unnamed protein product; ORF 1"  
 /codon\_start=1  
 /protein\_id="CAA38620.1"  
 /db\_xref="GI:4868"  
 /db\_xref="GOA:Q04318"  
 /db\_xref="SPTREMBL:Q04318"  
 /translation="MEMDRLFSAYSGDYKTYISGENSESELEVAICIKNGCFINIK"

## mRNA

mRNA  
 558..51279  
 /note="ORF 1"  
 578..1279  
 /note="unnamed protein product; ORF 1"  
 /codon\_start=1  
 /protein\_id="CAA38620.1"  
 /db\_xref="GI:4868"  
 /db\_xref="GOA:Q04318"  
 /db\_xref="SPTREMBL:Q04318"  
 /translation="MEMDRLFSAYSGDYKTYISGENSESELEVAICIKNGCFINIK"

## CDS

CDS  
 558..51279  
 /note="ORF 1"  
 578..1279  
 /note="unnamed protein product; ORF 1"  
 /codon\_start=1  
 /protein\_id="CAA38620.1"  
 /db\_xref="GI:4868"  
 /db\_xref="GOA:Q04318"  
 /db\_xref="SPTREMBL:Q04318"  
 /translation="MEMDRLFSAYSGDYKTYISGENSESELEVAICIKNGCFINIK"

FPFETLKRYSHRPPDLGVAVKFISKVINDYLLIEKYPFVTVVTPSCSVESYLL  
 ENFKELYNVKAIMLYNDLYFSDPEPELFLVSAFVNKNEVSDQSKAKBEG  
 YFYKLDIDYSELSLEPKENIETDGFDPNTVIGENELCEEYERKFFRSRSGKIE  
 GVKELDEVVRMGIGI  
 mRNA  
 581..>1279  
 /note="ORF 1"  
 complement (<1276..4278)  
 /note="ORF 2"  
 complement (1276..4275)  
 /note="unnamed protein product; ORF 2, has similarity to  
 DNA polymerase"  
 CDS  
 /codon\_start=1  
 /protein\_id="CAA38621.1"  
 /db\_xref="GI:4870"  
 /db\_xref="GOA:Q09038"  
 /db\_xref="SPTREMBL:Q09038"  
 /translation="MKLEFLKNEVSYTETDLDFLKSEVFKIHPYTSIEKLEKNS  
 TENSIRSTIYVNGQEIYKSLNGLIEEAKQAMKRFIQLNRNRPVLPKRKKK  
 LKPDHDETIIHKYKIDNLGNMDETQFTHKDSIIIPKRSILKHIHVRKYNY  
 YCVTLTLGFTTIGGVTSKVPILVCGIPNNYDERMTKALNDIERIPDEEKESN  
 IVCTRILITITQAIWKSIDNWPYKTRKFKKPDFTLFLASNEENCLQCCEKLGK  
 WYNNKLEMLPQKIIITYIPALMDIQVNSMEDLTDYNEEYISLDCKRIALLKWN  
 GHGIVITLPTKIQKIQKQLEVEDTNEIFEEDIESFTNEKQFPIPLICWSKT  
 NGSIHKIGKNCINDFVQYIINLKNVILYAFSGVDYQHVLPYFKKCIKDKYIHK  
 YHMTYSELFKPSKILKDPFLILTLSDKAAKAFVINKSGFPHSLKIDWDLQKI  
 YPHWIKTORKILREPHDKNLKINTTLEEDNITNNSITKEAIEVCSIDVIAKQVM  
 QKFKLVENLQISIEKTTLSQLSMKIMEASLDRIELVYPTSEYQPIRNSIYGG  
 RVAANGTYNEEIVADVSLPSAMKLLSHSGKSKVPSINPKHGIYDVTLIHS  
 NSEPHYLEFVPRIDKILMSWFERHGTTHYDILLALIQGFEIICHSGIYPNKG  
 YVNFNFIDKGLYTDIHTNCKEQPCIPRMVAKIALNGGQYKGVOKPIDKEVYIVK  
 RDVAGCENLSQNEGEICIGKIVKPKPFNLGDGEDYDKMIIERDDDDPIYTCQGV  
 SILSARVRLVNLCKEGLEIIVSDTDSIFVRKNTDWBELFKNTGKNIJELSDID  
 KTNALITHYMLIGGPVAFQVKNKNIITKLHCKGVPTEMLSIDQFELLHKKDKK  
 LAYFEILRLKLVTVTDKILDKIQI"  
 complement (<4283..5995)  
 /note="ORF 3"  
 complement (4283..5977)  
 /note="unnamed protein product; ORF 3"  
 /codon\_start=1  
 /protein\_id="CAA38622.1"  
 /db\_xref="GI:4871"  
 /db\_xref="GOA:Q09048"  
 /db\_xref="SPTREMBL:Q09048"  
 /translation="MFKKFLANQLSEPLVNIIELEARMKIQQDIDIPQIEITITY  
 RSTWPSIVFRNLNSELNSELVEKIPNDFIALSIEQSYNSEKFLPESNTRK  
 IYVRKLODPIVEITKNNENLEVEFDINNEVEKILKEMKSPYPPVKPMEISS  
 NLAKLSNDWCISPKADGSHIIQYINDESSIIHDNGFTSIDGNEKNIIIPRYVY  
 EALMNNKILYFDCLAYDKNIPKLDVIKREVLKKEKTEILFNNVALKSYLKP  
 QIKTDGYIISNKKKLYKSKFINTVDLRYNGYLLLENEKPSDRIPKDTYNFED  
 KIYEPDIEMNLIKERKORTIANYRFPYDDNPYIKIINGIGIPTLYRHNNKIKTELLNI  
 LPKSTLLDIGSGKGGDIHKVHLKPKIYAVDPILDRLQHSKNVVEIRDYVQNIYQLI  
 DYNYSIILFVPWDDIYLDVIOKAKHVIVAMNSPNKPYECDSFKCKVTDKVELNIPHT  
 TSNFVIENKTVYKILKNLKGNGWYHQBIKIMETIGTIOERELSRMYTHYLLKKIK  
 6166..>7927  
 /note="ORF 4"  
 6170..7927  
 /note="unnamed protein product; ORF 4"  
 /codon\_start=1  
 /protein\_id="CAA38623.1"  
 /db\_xref="GI:4872"  
 /db\_xref="GOA:Q04332"  
 /db\_xref="SPTREMBL:Q04332"  
 /translation="WNLFTVSDIKIEQYGFVNTKQNFIIISYLHPLNYPYSLVICYE  
 VGLGKTYASCLAHVLYSGYKVLVLSLSSIASISNFRNEYNKVITDSLTSEKNIDV  
 GSKFSFYNSKENNEEGLVIDEVNRLRELAIRYTIKNGDSDMSINKILVITAT  
 PMIDSKELDSINTNEKTIIFSENKISKDKINYGSNINGITELFLSMKQGLE  
 DYKSLKANDVTYTRQASISMNDEFNPEIPLDEQSKISRVLETLDGELTVIFC  
 FVYKGLDMKAVLSHFQNGPLNGSKVAVIDGRTSHKDVKEILKVFNSIANTNG  
 KOIQLVIGSSSESTLIRIRHLHLSFPWYGVQKOSIGRVNIGSHGLDKLTIN  
 VILHAYAEENFKGDIIDWETAVRKVEIDKDLKIENKVENIVEKISVPEVDNN  
 LVINKWVWDLRNCFQNKFKISWCRVYKDKAIGNIBNTKIIGNIPDKIKNTPL  
 KNGYTVRSICDNKRLSLYIDBEKVNKFTKRGKLLANVINIAKLNGCESTSDIINT  
 LCKENRYFDKIEYDLQY"  
 mRNA  
 CDS

7832..>8387  
 /note="ORF 5"  
 7873..>8387  
 /note="ORF 5"  
 7908..8387  
 /note="unnamed protein product; ORF 5"  
 /codon\_start=1  
 /protein\_id="CAA38624.1"  
 /db\_xref="GI:4873"  
 /db\_xref="GOA:Q04306"  
 /db\_xref="SPTREMBL:Q04306"  
 /translation="MIYNNIDTDLISLGKAKKFTTEEEKWITPIYVYGESLDFLKNK  
 YKINKIENSYGKGFVTKISKEYKSIIESKLGTVSPISDGSFRAFNNKTKIN  
 EKNLDDVSFNACVSLVFTIYKTEKTKLQIYIKDLVVVEIKDELEIFELKLSLAM"  
 8401..>11365  
 /note="ORF 6"  
 8408..11365  
 /note="unnamed protein product; ORF 6, has similarity to  
 RNA polymerase"  
 /codon\_start=1  
 /protein\_id="CAA38625.1"  
 /db\_xref="GI:4874"  
 /db\_xref="GOA:Q04387"  
 /db\_xref="SPTREMBL:Q04387"  
 /translation="MEYTDIGQIKIYETLPEDLVSTIKNTPISCIEMKLSVCGIYVNN  
 RELPMVSGKFDLHNDILIGIYFIIDGICSVNNIKMTERISFSKORAYITDNSKIE  
 IKDMFNYIYKDNKSRKMSLPINWRDIIKYSNKELEDHLSLIYTKSDKVIKNEIDL  
 MIICVPMFCWILGIGIKKESVHPWRLITPGELIYDIINKGDEVVVKCFTRKVIYNNVN  
 VYVLCCKHVNIVGDIIESIRRTIPTQRENTNENMVKLEDKYKICPVQTSDGALCGT  
 YIFNRYGVIKNSIVSVTASLIPYKKNPIRSMFSCKMMKQALSGENIYSKNLNT  
 KSLIFGGHGDIVVAIMFWYGNIEDAIVINKTASRPSKRCIYRENNCKIIGIYV  
 KXNIIYKQGLLYTNPVELVKTIBFIYSDQGTIDELKHSKYLKLVILKERDLEVG  
 DMTSRHGGQGVVSLIANNENFYKMKKEIDLLINHPAPPMTMGQIKEMGN  
 EXEVYNNKLDNKLIVRCFCYALRHQVDDKIQYRNSGNDIINKQPVSGKXNGGL  
 RFGOMERDILIGLAWNTLKWSDIKTNIYVCPKSGINKPCCKIKRESHQYFKICL  
 SYWEALGHDLIKNNKYSIIBPDSDLKKTTLKFGDLDPLDIRYKIVLPCLRS  
 TYNLKLYHLDNTKEIKELKLLKKGAVHKLVEGRVHDCHRSIVPDPPTLDIY  
 TVKIPUGANISTSYGLARQPSLNVDSMKIYLIQGNKTIISFNLCKSNADFDGD  
 EMNIYKNEESIQLKQKLVREKTDQYIILGKELIINKGLTANQGGIKLEKSGK  
 KEFNTHMFEKIGQVTINENIKYTIQPYNGEIKDDDWIYLSMAARQSAASIGVNTPI  
 TGVLESVCNQMYI"  
 complement (<11371..11777)  
 /note="ORF 7"  
 complement (11371..11769)  
 /note="unnamed protein product; ORF 7"  
 /codon\_start=1  
 /protein\_id="CAA38626.1"  
 /db\_xref="GI:4875"  
 /db\_xref="GOA:Q09059"  
 /db\_xref="SPTREMBL:Q09059"  
 /translation="WNEIDILNNLYFSEVTORTLSAFHTFGSDASLIDEINSIKRPFIT  
 NAKVNKLVKRGREDGKFRYFSDNSCDIVDLEDLSKPAYENKYDFSVIERKALSQNI  
 NPTSSLUKLSGHEIKKIKKEPSEIIFFM"  
 complement (<11371..11761)  
 /note="ORF 7"  
 complement (<11766..12027)  
 /note="ORF 8"  
 complement (<11766..12011)  
 /note="ORF 8"  
 complement (11766..12011)  
 /note="unnamed protein product; ORF 8"  
 /codon\_start=1  
 /protein\_id="CAA38627.1"  
 /db\_xref="GI:4876"  
 /db\_xref="GOA:Q04337"  
 /db\_xref="SPTREMBL:Q04337"  
 /translation="WTAINVDLELYSYNDIPKEILSSVYLSGKCNKSGKIKLTLKAPG  
 NEFIQGGPYILPPIRLMDSFEYVAFEVIVCDVDYVEK"  
 11989..>13389  
 /note="ORF 9"  
 12013..13389  
 /note="unnamed protein product; ORF 9"

mRNA  
CDS

mRNA  
mRNA  
mRNA  
CDS

mRNA  
CDS

```
/codon_start=1
/protein_id="CAA38628.1"
/db_xref="GI:4877"
/db_xref="GOA:004331"
/db_xref="SPTREMBL:004331"
/translation="HEIISYKIDFIYDKINEVWVSDNFQISEWEILRLKKTRTLQKS
YMDRSPYBPTQYTIALYSQNKVCLHNGKNVISPKNDEIPNLKLIKSTFTPLP
TYIKQDFHVMITLKNDEIKVLEFKRPSLTEGIILYNKMKINYNLSFSYKGENQ
KEPKLFCINTKMYEKDSKIKLXSKISFSDTHKSIDFRWYKAKHKNQOEWPKIS
KQLENSIEFPKNSIWIYYPQKFGVLARNDDNPLQIPIKLYKDHKSKNKNLHY
LQTKPKNDITIKIPIHLRAKGLKVAQIKLITNEKYIEMSYNGIIEITDKNPEYHY
ENILAKTKIKINIDLDADLLANONNERIMILKNNKWKISQSGPRIKNIITIPWYKQIL
HDYKILRLKXGKIIDLISNIGIVDENADIIITWETENLYSSKLTQLRLITFKPKQK
```

Query Match 10.2%; Score 50.6; DB 8; Length 14281;  
Best Local Similarity 48.8%; Pred. No. 0.099;  
Matches 137; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 195 TTGAGCTAGTCGATTATTGGATGATCCATTCTTCATCGTTTTTCTTCTGCTCGAAGT 254  
DB 11290 TTGCTTGGCGCCATAGATAGTATACCAATCATCTTTTATTCGCCGTGTGAAG 11231

QY 255 TCTGTATAACACGATTGCTGTGCGGATGTGCTATACCTAGCGGTATCGAGAACTA 314  
DB 11230 TTGATCTATATAAGTTTTATGTTTCGTTTATAGTACTTGCACCTATTTTTCGAACAT 11171

QY 315 GGGTTTCGAGTCAATTTGCCCTTTTGGTTATATCTGCTCGATAACGATTCATCTGG 374  
DB 11170 ATGTTATAATTAATCTTTGCTTTAGATCTCTTTTCATCATTAATTTATGCTTTG 11111

QY 375 ATTAGGGTTTTTAAGTGTGACGTTTATGATTTCCAAATTTCTTCAAAATTTAGTTATGATA 434  
DB 11110 TCTATTTGGGTCAATCTTTATTTATTTCTTTCTTCTTAAATATAGTCTTGATTT 11051

QY 435 ATGMAATCCGAAATGACGTGTTCAATTTCTGTTAAATCC 475  
DB 11050 ATCTTCTTAACTAATTAATTTTGTTTTAAATCTGTATAC 11010

RESULT 10  
LOCUS I66494 7218 bp DNA linear PAT 28-DEC-1997  
DEFINITION Sequence 14 from patent US 5670367.  
ACCESSION I66494  
VERSION I66494.1 GI:2724471  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (Bases 1 to 7218)  
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.  
TITLE Recombinant fowlpox virus  
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;  
FEATURES Location/Qualifiers  
source 1..7218  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN  
Query Match 10.0%; Score 49.6; DB 6; Length 7218;  
Best Local Similarity 6.1%; Pred. No. 0.18;  
Matches 25; Conservative 213; Mismatches 172; Indels 0; Gaps 0;

QY 14 CAGGTATGATTGTTTGTAAATAGATCGGGTTAGTCTTCCATCTTTTAAATGT 73  
DB 1046 CAGGTGAGGAGCTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1105

QY 74 TTTTCTGTACTGTCTCGGATCTGATTTTACCAATAGATTTTCGGGTTTTGTC 133  
DB 1106 YY 1165

QY 134 ATTCCAGTTTGAATAAATACGTCCTCTTTTAAAGTTTCTGCGATCGATAACCTGTGAAG 193  
DB 54643 GTTTTITAGTAAAGTTTTTTTCTGTTTAAATAAAGGAGATTAAGAGTACG 54702

Db 1166 YY 1225

QY 194 ATTGAGCTAGTCGATTATTGATGATCCATTCTTCATCGTTTTTCTTCTGTCGAAG 253  
DB 1226 YY 1285

QY 254 TTTCTGTATAACGATTTGCTGTGTCGATTTGCTATTACTACCGCTGATCGAGAAT 313  
DB 1286 YY 1345

QY 314 AGGGTTTTCGAGTCAATTTTCCCTTTTGGTTATATCTGCTCGATAACGATTCATCTG 373  
DB 1346 YY 1405

QY 374 GATTAGGGTTTAAAGTGGTGACGTTTAGTATTCCAAATTTCTTCAAAATTT 423  
DB 1406 YYYGGTACCAATTTCTTATCTCT 1455

RESULT 11  
LOCUS AX344573 349980 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 24 from Patent WO0200932.  
ACCESSION AX344573  
VERSION AX344573.1 GI:18492459  
KEYWORDS synthetic construct  
SOURCE synthetic construct  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of known genetic parameters within the mhc  
JOURNAL Patent: WO 0200932-A 24 03-JAN-2002;  
EpiGenomics AG (DE)  
FEATURES Location/Qualifiers  
source 1..349980  
/organism="synthetic construct"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)-Original length of seq 1: 3.673778 <223>-split as follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 <223>-Original length of seq 2: 3.673778 <223>-split as follows.-seq 14 0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16 600.001 949.980-seq 17 900.001 1.249.980-seq 18 1.200.001 1.549.980-seq 19 1.500.001 1.849.980-seq 20 1.800.001 2.149.980-seq 21 2.100.001 2.449.980-seq 22 2.400.001 2.749.980-seq 23 2.700.001 3.049.980-seq 24 3.000.001 3.349.980-seq 25 3.300.001 3.649.980-seq 26 3.600.001 3.673.778"

ORIGIN  
Query Match 10.0%; Score 49.6; DB 6; Length 349980;  
Best Local Similarity 49.6%; Pred. No. 0.13;  
Matches 127; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 18 TATGATTCTTGTAAATAGATCAGGGTTTAGTCTTTCATCTACTTTTAAATGTTT 77  
DB 54523 TGTGGTTTAAATGTTATAGATAGTTTTTTCAGATTTTGTAGTTTTTTTTTT 54582

QY 78 TCTGTACTGTCTCGCATCTGATTTTACGCAATAGATTTCCGGTTTCTGCCATC 137  
DB 54583 TTTTGTAGTAATTTTATATTTTTTTTTTAAAGTTTTTTAAAGAAATTTTATTT 54642

QY 138 CAGTTTGAATAAAGCTCCGCTTTTAAAGTTTGTGATCGATAAAGCTGTGAAGATTG 197  
DB 54643 GTTTTITAGTAAAGTTTTTTTCTGTTTAAATAAAGGAGATTAAGAGTACG 54702







Qy 403 ATTCCAATTCTTCAAAAATTAGTTATGGATAATGAAAAATCCGAAATGACTGTTCAT 462  
Db 129446 TTKRTWWTYYTTTWTTKKTAAWAAWKKWWWWYCWAAWTTKGRWKTCTCWWAAAAGKT 129505  
Qy 463 TCTTGTTAAATGGCGAGATCC 483  
Db 129506 TYCMRCTAWTTWYATCTTC 129526

Search completed: June 20, 2004, 03:45:49  
Job time : 3467.42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2004, 01:40:43 ; Search time 59.125 Seconds  
(without alignments)  
3923.374 Million cell updates/sec

Title: US-09-000-062-6  
Perfect score: 418  
Sequence: 1 tgagtagattcttcgattc.....taattgtgaacagatccc 418

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 418   | 100.0       | 418    | 4  | US-09-000-062-6     |
| 2          | 43.8  | 10.5        | 5852   | 1  | US-07-867-106-2     |
| 3          | 42.4  | 10.1        | 8537   | 4  | US-10-204-708-41    |
| 4          | 40.2  | 9.6         | 5455   | 4  | US-10-204-708-34    |
| 5          | 40.2  | 9.6         | 11049  | 4  | US-10-204-708-23    |
| 6          | 39.6  | 9.5         | 11015  | 4  | US-10-204-708-55    |
| 7          | 39.4  | 9.4         | 5340   | 4  | US-09-627-122-21    |
| 8          | 39.2  | 9.4         | 11050  | 4  | US-10-204-708-86    |
| 9          | 39    | 9.3         | 5501   | 4  | US-10-204-708-37    |
| 10         | 38.4  | 9.2         | 5562   | 4  | US-10-204-708-63    |
| 11         | 38.2  | 9.1         | 5501   | 4  | US-10-204-708-38    |
| 12         | 38.2  | 9.1         | 6801   | 4  | US-10-204-708-61    |
| 13         | 38.2  | 9.1         | 6886   | 4  | US-10-204-708-20    |
| 14         | 38    | 9.1         | 786431 | 4  | US-09-751-389-3     |
| 15         | 37.8  | 9.0         | 5455   | 4  | US-10-204-708-33    |
| 16         | 37.8  | 9.0         | 6801   | 4  | US-10-204-708-62    |
| 17         | 37.8  | 9.0         | 19124  | 2  | US-08-487-826B-13   |
| 18         | 37.6  | 9.0         | 11050  | 4  | US-10-204-708-85    |
| 19         | 37.6  | 9.0         | 11811  | 3  | US-09-078-294-7     |
| 20         | 37.4  | 8.9         | 5686   | 4  | US-10-204-708-29    |
| 21         | 37.4  | 8.9         | 6306   | 4  | US-10-204-708-50    |
| 22         | 37.4  | 8.9         | 23439  | 4  | US-08-956-171E-38   |
| 23         | 37.2  | 8.9         | 832    | 4  | US-09-621-976-2813  |
| 24         | 37    | 8.9         | 6317   | 4  | US-10-204-708-11    |
| 25         | 37    | 8.9         | 10640  | 4  | US-09-417-485D-5    |
| 26         | 36.6  | 8.8         | 984    | 4  | US-09-134-001C-2100 |
| 27         | 36.6  | 8.8         | 19513  | 4  | US-10-204-708-40    |

|   |    |      |     |        |   |                     |                    |
|---|----|------|-----|--------|---|---------------------|--------------------|
| c | 28 | 36.6 | 8.8 | 640681 | 4 | US-09-730-988-1     | Sequence 1, Appli  |
|   | 29 | 36.4 | 8.7 | 6040   | 4 | US-10-204-708-70    | Sequence 70, Appl  |
|   | 30 | 36.4 | 8.7 | 6182   | 4 | US-10-204-708-88    | Sequence 88, Appl  |
|   | 31 | 36.4 | 8.7 | 8607   | 4 | US-10-204-708-71    | Sequence 71, Appl  |
|   | 32 | 36.4 | 8.7 | 202001 | 4 | US-09-734-674-3     | Sequence 3, Appli  |
|   | 33 | 36.2 | 8.7 | 6113   | 4 | US-10-204-708-14    | Sequence 14, Appl  |
|   | 34 | 36.2 | 8.7 | 7218   | 1 | US-08-232-453-14    | Sequence 14, Appl  |
|   | 35 | 36.2 | 8.7 | 8093   | 4 | US-10-204-708-32    | Sequence 32, Appl  |
|   | 36 | 36.2 | 8.7 | 11015  | 4 | US-10-204-708-56    | Sequence 56, Appl  |
|   | 37 | 36   | 8.6 | 6070   | 4 | US-10-204-708-10    | Sequence 10, Appl  |
|   | 38 | 36   | 8.6 | 19233  | 4 | US-10-204-708-45    | Sequence 45, Appl  |
| c | 39 | 35.8 | 8.6 | 2103   | 4 | US-09-543-681A-797  | Sequence 797, Appl |
|   | 40 | 35.8 | 8.6 | 5666   | 4 | US-10-204-708-30    | Sequence 30, Appl  |
|   | 41 | 35.8 | 8.6 | 8607   | 4 | US-10-204-708-72    | Sequence 72, Appl  |
|   | 42 | 35.4 | 8.5 | 522    | 4 | US-09-134-001C-1017 | Sequence 1017, Ap  |
|   | 43 | 35.4 | 8.5 | 8961   | 4 | US-10-204-708-80    | Sequence 80, Appl  |
| c | 44 | 35.2 | 8.4 | 6243   | 2 | US-09-056-075-1     | Sequence 1, Appli  |
|   | 45 | 35.2 | 8.4 | 8537   | 4 | US-10-204-708-42    | Sequence 42, Appl  |

ALIGNMENTS

RESULT 1  
US-09-000-062-6  
; Sequence 6, Application US/09000062  
; Patent No. 6338961  
; GENERAL INFORMATION:  
; APPLICANT: DEROSE, Richard  
; APPLICANT: CHAUBET, Nicole  
; TITLE OF INVENTION: ISOLATED DNA SEQUENCE CAPABLE OF SERVING AS REGULATORY  
; TITLE OF INVENTION: ELEMENT IN A CHIMERIC GENE WHICH CAN BE USED FOR THE  
; FILE OF INVENTION: TRANSFORMATION OF PLANTS  
; FILE REFERENCE: 022650-453  
; CURRENT APPLICATION NUMBER: US/09/000,062  
; CURRENT FILING DATE: 1998-05-29  
; EARLIER APPLICATION NUMBER: PCT/FR96/01109  
; EARLIER FILING DATE: 1996-07-17  
; EARLIER APPLICATION NUMBER: FR 95/08980  
; EARLIER FILING DATE: 1995-07-19  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 418  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-09-000-062-6

|                       |              |  |               |             |
|-----------------------|--------------|--|---------------|-------------|
| Query Match           | 100.0%       | Score 418;   | DB 4;         | Length 418; |
| Best Local Similarity | 100.0%       | Pred. No. 1.8e-102;  | Mismatches 0; | Indels 0;   |
| Matches 418;          | Conservative | 0;   |               |             |
| Qy                    | 1            | TCAGTACGATCTTCGATCCTCTTTGATTTCTCGGAAATATTTTCGGTGCATCGTGA   | 60            |             |
| Db                    | 1            | TCAGTACGATCTTCGATCCTCTTTGATTTCTCGGAAATATTTTCGGTGCATCGTGA   | 60            |             |
| Qy                    | 61           | AACACTAGGATCGCTCGATAGTGGTACGAAATAGCGAGATTAGTTCTTATCTTGG    | 120           |             |
| Db                    | 61           | AACACTAGGATCGCTCGATAGTGGTACGAAATAGCGAGATTAGTTCTTATCTTGG    | 120           |             |
| Qy                    | 121          | CCATTATCTTCTTCGCGCAATGATCTTCGATATAAGATTTTAGGTTAGAGATGAA    | 180           |             |
| Db                    | 121          | CCATTATCTTCTTCGCGCAATGATCTTCGATATAAGATTTTAGGTTAGAGATGAA    | 180           |             |
| Qy                    | 181          | TCGTATAGCTAGATTTCATCACCAGATAGTTCTTGTCTAGAAATCTCTGAAATCTCGA | 240           |             |
| Db                    | 181          | TCGTATAGCTAGATTTCATCACCAGATAGTTCTTGTCTAGAAATCTCTGAAATCTCGA | 240           |             |
| Qy                    | 241          | TAGTTTTCACATGTGTAATAGATTGTTCTTATTCGCGGATTTGTTAGTGGTTTTGAT  | 300           |             |
| Db                    | 241          | TAGTTTTCACATGTGTAATAGATTGTTCTTATTCGCGGATTTGTTAGTGGTTTTGAT  | 300           |             |





Matches 78; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
QY 209 GTTCTTTGCTAGATCTCTGAAATCTCGATAGTTTTCATGCTGTAATAGATTGTT 268  
DB 3003 GTTTTGTGTTTATTTATTTGTTATGCTGTTTATAAAGTTTATAAGAGAAAGTT 3062  
QY 269 CTTATTCCGCGATTGTTGATTAGGGTTTTCATTTCTTGATTATGCGATTGCAATTAGG 328  
DB 3063 TGTATATGTTGTTGTTTAGAGAGTCTTTTATTAATTGATTGGTTAGGAGTAATATAG 3122  
QY 329 ATTCTTTGCTTTGTTGTTGA 350  
DB 3123 GTTTTGTGAATGCTGTTTA 3144

## RESULT 7

US-09-627-122-21  
; Sequence 21, Application US/09627122  
; Patent No. 6472521  
; GENERAL INFORMATION:  
; APPLICANT: Uhlmann, Eugen  
; APPLICANT: Greiner, Beate  
; APPLICANT: Unger, Eberhard  
; APPLICANT: Gothe, Gislinde  
; APPLICANT: Schwardel, Marc  
; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE INHIBITION OF HUMAN egs  
; FILE REFERENCE: 02481.1678  
; CURRENT APPLICATION NUMBER: US/09/627,122  
; CURRENT FILING DATE: 2000-07-27  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 5340  
; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
US-09-627-122-21

Query Match 9.4%; Score 39.4; DB 4; Length 5340;  
Best Local Similarity 51.3%; Pred. No. 0.2; Indels 5; Gaps 2;  
Matches 143; Conservative 0; Mismatches 131; Indels 5; Gaps 2;  
QY 86 GTACGAATTAGCGAGATTAGTTCTATCTTGGCCATATCTTCTTCCTTCGCGCAAT 145  
DB 2960 GCATTATATTAGACATATTTGTTTAAATTTGTTGAAAGATTTTTCGATTATAAAT 3019  
QY 146 GATCTTCGGTATAAGATTTAGTTAGTATGAGATGAATCGTATAGCTAGATTTTCATCACCAG 205  
DB 3020 TTCTTCTTAAAGAATGATATTTATTTAATAGAAATGCTGTAATATATTTCTT---CTA 3076  
QY 206 ATAGTTTCTTTGCTAGAAATCTGAAATCTCGATAGTTTTCACATGCTGTAATAGATT 265  
DB 3077 TTAATTTTAAATAATATTGATGAATGCTTGATATTTTATATTTTGAATAGTAT 3136  
QY 266 GTTCTATTCCGGGATGTTGATTAGGGTTTTCATTTGATTTCTCGATTATGGATGCAATTA 325  
DB 3137 CTTTAAAGAAAGAAATATTT---TTATTTTGTAGATCATCTGTATTATCCATTTTATTA 3194  
QY 326 GCGATTTCTTTGTTGTTGTTGATCTTACGATACATT 364  
DB 3195 TAAATTTTATTTTATTTTACTTTTTCATTAATAATTT 3233

## RESULT 8

US-10-204-708-86  
; Sequence 86, Application US/10204708  
; Patent No. 6677731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; TITLE OF INVENTION: by Assessing DNA Methylation

; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 86  
; LENGTH: 11050  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-86  
Query Match 9.4%; Score 39.2; DB 4; Length 11050;  
Best Local Similarity 46.9%; Pred. No. 0.27; Indels 0; Gaps 0;  
Matches 122; Conservative 0; Mismatches 138; Indels 0; Gaps 0;  
QY 154 GTATPAAAGATTTTAGTTAGAGATGAATCGTAGCTAGATTTTCATCACCAGATAGTTTC 213  
DB 268 GTAAATATTCGATAAATGAGTGAATTTTATTTTATTTAGTATTTTATGATATGGGTTTG 327  
QY 214 TTTGCTAGATCTCTGAAATCTCGATAGTTTTCATGATGTTTCAATGTAATAGATTGTTCTAT 273  
DB 328 TATGACGTGTTTGGTAGATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTAT 387  
QY 274 TCGCGATTTGTTGATTAGGGTTTGTATTTCTTGATTCGATTCGATTCGAAATAGGATTTT 333  
DB 388 TGAGTTTAAATATATTTGTTTAGTTTGTATATTAATTTTGTAGGTTTTC 447  
QY 334 CTTTGGTTTGTGTTGATCTAGATACATCTCGAATGAAATAGATGATGATCTAAAT 393  
DB 448 AATTTGGTTGTTTTTTTATTTTGAATATTTTTTTTGGAGTTTATATGTTGTTT 507  
QY 394 CTTGTTAAATTTGTGAACAG 413  
DB 508 TTAGTTTATTTAGTTTATAG 527

## RESULT 9

US-10-204-708-37  
; Sequence 37, Application US/10204708  
; Patent No. 6677731  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; TITLE OF INVENTION: by Assessing DNA Methylation  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204,708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 98  
; SEQ ID NO 37  
; LENGTH: 5501

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-37

```

| Query Match           | 9.3%;           | Score 39;  | DB 4;     | Length 5501; |
|-----------------------|-----------------|--|-----------|--------------|
| Best Local Similarity | 49.8%;          | Fred. No. 0.86;  |           |              |
| Matches 126;          | Conservative 0; | Mismatches 125;  | Indels 2; | Gaps 1;      |
| QY                    | 102             | GATTAGTCTTCTATCTTGGCCATTATCTTCTTCTTCGCGCAATGATCTTCGGTATAAAG    | 161       |              |
| Db                    | 1246            | GTTGATGTAATATTTTAAAAATTTTGGGGTTTTTGTGTAATGTTATGCAGTGTAAATG     | 1305      |              |
| QY                    | 162             | ATTTTAGGTTAGAGATGAATCGTATAGCTAGATTCATCACCAGATAGTTCTTCGTCTA     | 221       |              |
| Db                    | 1306            | TTATTAGATAAGAGGGTTTTTTTAAATTTTTTGGGATAA--TTTATTATTTTATTTTGA    | 1363      |              |
| QY                    | 222             | GAATCTCTGAAATTCGATPAGTTTTACATGTTGTAATAATAGATTGTTCTTATTTCGGCGAT | 281       |              |
| Db                    | 1364            | TTTTTTATTGAGATGGTTGAATGAGTTTAAAGATTTATGATTTTTATTTTTTCGTGTTATTA | 1423      |              |
| QY                    | 282             | TGTTGATTAGGGTTTTGATTTCTTCGATTATCGGATTGCAATTAGGGATTTCTTTGGTGT   | 341       |              |
| Db                    | 1424            | TATTTAGTATATTTTTTTTATTTTATTTTGTAAAGAGATCGGGTTTCGTTATGTTGTTAGTT | 1483      |              |
| QY                    | 342             | TTGTGTGTGATCTT   | 354       |              |
| Db                    | 1484            | TGGTTTTGAAATT  | 1496      |              |

```

RESULT 10
US-10-204-708-63
; Sequence 63, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 63
; LENGTH: 5562
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-63

```

|    | Query Match           | 9.2%;  | Score 38.4;     | DB 4;     | Length 5562; |
|----|-----------------------|--|-----------------|-----------|--------------|
|    | Best Local Similarity | 49.5%;   | Pred. No. 0.37; |           |              |
|    | Matches 99;           | Conservative 0;  | Mismatches 101; | Indels 0; | Gaps 0;      |
| QY | 208                   | AGTTTCCTTCCTAGAAATCTCTGAAATTCGATAGTTTTCCACATGTTGAATAGATTGT   | 267             |           |              |
| Db | 367                   | AGTTTTTAGGTTTTTTTTTAATTATAAGTTGGTTGGTGATTTTCGATTTATTTT       | 426             |           |              |
| QY | 268                   | TCTATTTCGGCGATGTTTGANTAGGGTTTTGATTTTCTTGATTATCGGATTCGAATTAGG | 327             |           |              |

|    |     |   |     |
|----|-----|---|-----|
| Db | 427 | TTTTTATAGTTTTTTCGATGAGTGTTTTTTTTGATTAAAGCGGTATTTTT            | 486 |
| Qy | 328 | GATTTCTTTGGTTTTGTCTGATCTTAGATACATCTCTGCAATGAATACGTATGGAT      | 387 |
| Db | 487 | TTATTAGGTTTTTATATTTTTTTTTTTTGTGATTGAAGAGTTTTTTTAGTTTGTGATAAAT | 546 |
| Qy | 388 | CTAAATCTTTGTTAATTGGT  | 407 |
| Db | 547 | ATTTTTGTTTTTTATTGGTT  | 566 |

```

RESULT 11
US-10-204-708-38
; Sequence 38, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 38
; LENGTH: 5501
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-38

```

| Query Match           | 9.1%;  | Score 38.2;   | DB 4;          | Length 5501;      |
|-----------------------|--------|---|----------------|-------------------|
| Best Local Similarity | 49.7%; | Pred. No. 0.42;   |                |                   |
| Matches               | 97;    | Conservative 0;   | Mismatches 98; | Indels 0; Gaps 0; |
| Qy                    | 210    | TTTCTTTGTCAGAACTCTCGAAATCTCGATAGTATTACATGTGTAATAAGATTGTC      | 269            |                   |
|                       |        |   |                |                   |
| Db                    | 3630   | TGCTTTGTTAGTATGTTTTTATTGTATGTTTTTTTTTTGGTGTGTTTTTTT           | 3689           |                   |
|                       |        |   |                |                   |
| Qy                    | 270    | TTATTCGGCGATTTGTAATPAGGTTTGTGATTTTCTTGATATGCGAATTCGAATAGGGA   | 329            |                   |
|                       |        |   |                |                   |
| Db                    | 3690   | TTATTTAGGTTTTAGTTTGGTAGTTATATTTTGTACGTTTTTGGATTAGGTTTAGGTT    | 3749           |                   |
|                       |        |   |                |                   |
| Qy                    | 330    | TTTTCTTTTGGTTTTGTTGTCATCTTACGATACATTCCTGCAATTCGAATCGTATGGATCT | 389            |                   |
|                       |        |   |                |                   |
| Db                    | 3750   | TTTTATTGTGTGTTTTTTTTTATAGTATTACAGAGGATTGGAATATTATTACGGTGTGT   | 3809           |                   |
|                       |        |   |                |                   |
| Qy                    | 390    | AAATCTGTTTAAATTT  | 404            |                   |
|                       |        |   |                |                   |
| Db                    | 3810   | ATATTATGATTTTT  | 3824           |                   |
|                       |        |   |                |                   |

```

RESULT 12
US-10-204-708-61
; Sequence 61, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication

```

;; TITLE OF INVENTION: by Assessing DNA Methylation  
;; FILE REFERENCE: 5013.1012  
;; CURRENT APPLICATION NUMBER: US/10/204,708  
;; CURRENT FILING DATE: 2003-05-06  
;; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
;; PRIOR FILING DATE: 2001-04-06  
;; PRIOR APPLICATION NUMBER: DE 10019058.8  
;; PRIOR FILING DATE: 2000-04-06  
;; PRIOR APPLICATION NUMBER: DE 10019173.8  
;; PRIOR FILING DATE: 2000-04-07  
;; PRIOR APPLICATION NUMBER: DE 10032529.7  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: DE 10043826.1  
;; PRIOR FILING DATE: 2000-09-01  
;; NUMBER OF SEQ ID NOS: 98  
;; SEQ ID NO 61  
;; LENGTH: 6801  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-61

Query Match 9.1%; Score 38.2; DB 4; Length 6801;  
Best Local Similarity 50.8%; Pred. No. 0.45;  
Matches 91; Conservative 0; Mismatches 88; Indels 0; Gaps 0;  
QY 170 TTAGAGATGATCGTAGTAGTATTCATCACCAGATAGTTCTTGTCTAGAAATCTCT 229  
DB 192 TGAGGGATAAATGATTTTCGTGGTAGTGTGATTTATATTTTATTAATAATGAT 251  
QY 230 GAAATCTCGATAGTTTTCACATGCTGTAATAGATTTCTTATTCGGCGATTGTTGAT 289  
DB 252 GAGGGTTTAAATTTTATATTTTGTATATTTTATTTTATTTTATTTTATTTT 311  
QY 290 AGGGTTTGAATTTCTGATTAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 348  
DB 312 TTTTAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 370

RESULT 13  
US-10-204-708-20  
;; Sequence 20, Application US/10204708  
;; Patent No. 6677731  
;; GENERAL INFORMATION:  
;; APPLICANT: OLEK, Alexander  
;; APPLICANT: PIEPENBROCK, Christian  
;; APPLICANT: BERLIN, Kurt  
;; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
;; FILE REFERENCE: 5013.1012  
;; CURRENT APPLICATION NUMBER: US/10/204,708  
;; CURRENT FILING DATE: 2003-05-06  
;; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
;; PRIOR FILING DATE: 2001-04-06  
;; PRIOR APPLICATION NUMBER: DE 10019058.8  
;; PRIOR FILING DATE: 2000-04-06  
;; PRIOR APPLICATION NUMBER: DE 10019173.8  
;; PRIOR FILING DATE: 2000-04-07  
;; PRIOR APPLICATION NUMBER: DE 10032529.7  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: DE 10043826.1  
;; PRIOR FILING DATE: 2000-09-01  
;; NUMBER OF SEQ ID NOS: 98  
;; SEQ ID NO 20  
;; LENGTH: 6866  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-20

Query Match 9.1%; Score 38.2; DB 4; Length 6866;

Best Local Similarity 49.4%; Pred. No. 0.45;  
Matches 126; Conservative 0; Mismatches 128; Indels 1; Gaps 1;  
QY 155 TATAAAGATTTTAGGTTAGAGATCAATCGTAGTAGATTTTATCATCACCAGATAGTTCT 214  
DB 1005 TATTTAGATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGAT 1064  
QY 215 TTGTCTAGAAATCTGAAATTCGATAGTTT-CACATGCTGTAATAGATTTCTTCTTAT 273  
DB 1065 TTATTTAGAGTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGAT 1124  
QY 274 TCAGGATTTGATAGGTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTT 333  
DB 1125 TTGGGATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTT 1184  
QY 334 CTTTGGTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTT 393  
DB 1185 TTTTATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTT 1244  
QY 394 CTTGTTAAATTTGTTG 408  
DB 1245 TTATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTTTCGATATTTT 1259

RESULT 14  
US-09-751-389-3/c  
;; Sequence 3, Application US/09751389  
;; Patent No. 6630334  
;; GENERAL INFORMATION:  
;; APPLICANT: GUEGLER, Karl et al  
;; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
;; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
;; TITLE OF INVENTION: THEREOF  
;; FILE REFERENCE: CL001067  
;; CURRENT APPLICATION NUMBER: US/09/751,389  
;; CURRENT FILING DATE: 2001-01-02  
;; NUMBER OF SEQ ID NOS: 8  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 3  
;; LENGTH: 786431  
;; TYPE: DNA  
;; ORGANISM: Human  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (1)...(786431)  
;; OTHER INFORMATION: n = A,T,C or G  
US-09-751-389-3

Query Match 9.1%; Score 38; DB 4; Length 786431;  
Best Local Similarity 55.2%; Pred. No. 1.8;  
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 229 TGAATTTCTCGATAGTTTTCACATGCTGTAATAGATTTCTTATTCGGCGATTGTTGAT 288  
DB 455675 TGGAAATTCATTATATTTCCATCTCTATAATTTGTTACTATATCTTCATTTCTT 455616  
QY 289 TAGGGTTTTCGATATTTTCGATATTTTCGATATTTTCGATATTTTCGATATTTTCGAT 348  
DB 455615 CCCAGTTTTCGATATTTTCGATATTTTCGATATTTTCGATATTTTCGATATTTTCGAT 455556  
QY 349 GATCTTACGATACA 362  
DB 455555 ATCAAGAGATTA 455542

RESULT 15  
US-10-204-708-33  
;; Sequence 33, Application US/10204708  
;; Patent No. 6677731  
;; GENERAL INFORMATION:  
;; APPLICANT: OLEK, Alexander  
;; APPLICANT: PIEPENBROCK, Christian  
;; APPLICANT: BERLIN, Kurt

/ TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
/ TITLE OF INVENTION: By Assessing DNA Methylation  
/ FILE REFERENCE: 5013.1012  
/ CURRENT APPLICATION NUMBER: US/10/204,708  
/ CURRENT FILING DATE: 2003-05-06  
/ PRIOR APPLICATION NUMBER: PCT/EP01/03971  
/ PRIOR FILING DATE: 2001-04-06  
/ PRIOR APPLICATION NUMBER: DE 10019058.8  
/ PRIOR FILING DATE: 2000-04-06  
/ PRIOR APPLICATION NUMBER: DE 10019173.8  
/ PRIOR FILING DATE: 2000-04-07  
/ PRIOR APPLICATION NUMBER: DE 10032529.7  
/ PRIOR FILING DATE: 2000-06-30  
/ PRIOR APPLICATION NUMBER: DE 10043826.1  
/ PRIOR FILING DATE: 2000-09-01  
/ NUMBER OF SEQ ID NOS: 98  
/ SEQ ID NO 33  
/ LENGTH: 5455  
/ TYPE: DNA  
/ ORGANISM: Artificial Sequence  
/ FEATURE:  
/ OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-33

Query Match 9.0%; Score 37.8; DB 4; Length 5455;  
Best Local Similarity 53.8%; Pred. No. 0.54;  
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
QY 234 TTCTCGATAGTTTTCACATGTCGTAATAGATTCTTATTCGGCGATTGTTGATTAGG 293  
Db 3679 TTTTCTTTTGGGTTTTTTTATTTTATTTTGGGTTGGTAAATTAGT 3738  
QY 294 TTTTGATTTCCTGATTATGCGATTGCAATTAGGATTTCCTTGGTTTCTGTTGATCT 353  
Db 3739 TTTTAAATTTTATATTTTATAGATTGGAAGTGGTGTTCATATTTTAGAGTT 3798  
QY 354 TAGATACATTCCTGCAATTGAATA 378  
Db 3799 GTTATTAATTTTAAAGTTTAA 3823

Search completed: June 20, 2004, 05:03:40  
Job time : 61.125 secs